

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 07-08-03
Searcher: Beverly 24994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site
____ STIC
____ CM-1
____ Pre-S
Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
____ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ Other CGN



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 97680

To: **Phuong Bui**
Location: **CM1-9E12**
Art Unit: **1638**
Tuesday, July 08, 2003

Case Serial Number: 09/938294

From: **Beverly Shears**
Location: **Biotech-Chem Library**
CM1-1E05
Phone: **308-4994**

beverly.shears@uspto.gov

Search Notes

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 18:49:00 ; Search time 2990 Seconds

(without alignments)
16566.213 Million cell updates/sec

Title: US-09-938-294-2

Sequence: 1 atccggctcgagcttgatcc.....gagggagcttcaatccttagt 1702

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Genbank1.*

- 1: gb_pa.*
- 2: gb_ptg.*
- 3: gb_in.*
- 4: gb_cm.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_dr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_pa.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vt.*
- 30: em_ptg_hum.*
- 31: AC024609/c
- 32: em_ptg_in.*
- 33: em_ptg_other.*
- 34: em_ptg_mus.*
- 35: em_ptg_pln.*
- 36: em_ptg_rod.*
- 37: em_ptg_mam.*
- 38: em_ptg_vrt.*
- 39: em_ptgo_hum.*
- 40: em_ptgo_mus.*
- 41: em_ptgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match Length	ID	Description
C 1	282.6	16.6 90341	8 AC024609	AC024609 Arabidops
C 2	282.6	16.6 119942	8 AC0007797	AC0007797 Arabidops
C 3	253.4	14.9 100685	8 AC023754	AC023754 Arabidops
C 4	224.8	13.2 146081	8 AP001366	AP001366 Oryza sat
C 5	89.4	5.3 2557	8 AY091763	AY091763 Arabidops
C 6	87.8	5.2 75948	8 AC037424	AC037424 Arabidops
C 7	87.8	5.2 11737	8 F6D8	AC008016 Arabidops
C 8	65.8	3.9 147292	2 AC023010	AC023010 Homo sapi
C 9	64.4	3.8 104486	2 AC109537	AC109537 Rattus no
C 10	58.2	3.4 42273	8 NCBI2M19	AL669987 Neurospor
C 11	58.2	3.4 122151	8 NCBI2H20	AL669988 Neurospor
C 12	57	3.3 152607	2 AC126306	AC126306 Rattus no
C 13	56.4	3.3 35762	8 AB026653	AB026653 Arabidops
C 14	56	3.3 70311	8 AF128457	AF128457 Oryza sat
C 15	56	3.3 77605	8 AF119222	AF119222 Oryza sat
C 16	56	3.3 142852	8 AF161269	AF161269 Oryza sat
C 17	55.6	3.3 42844	8 AC069166	AC069166 Homo sapi
C 18	55.4	3.3 151771	2 AC128127	AC128127 Rattus no
C 19	53	3.1 23990	1 SC6B10	AL109661 Streptomy
C 20	53	3.1 55493	2 AC096794	AC096794 Rattus no
C 21	53	3.1 141025	2 AC123528	AC123528 Oryza sat
C 22	52.4	3.1 176399	2 AC127769	AC127769 Rattus no
C 23	52	3.1 956	11 PM2D128	AL684743 Penicilliu
C 24	52	3.1 141307	8 AC084763	AC084763 Oryza sat
C 25	51.8	3.0 159184	2 AC130633	AC130633 Rattus no
C 26	51.2	3.0 114741	2 AC111484	AC111484 Rattus no
C 27	51	3.0 136917	2 AP005097	AP005097 Oryza sat
C 28	50.8	3.0 79734	2 AC096238	AC096238 Rattus no
C 29	50.8	3.0 125888	2 AC128139	AC128139 Rattus no
C 30	50.8	3.0 128049	2 AC111232	AC111232 Rattus no
C 31	50.6	3.0 1560	14 HS21E	M29384 Herpes simp
C 32	50.6	3.0 87500	8 AP004308	AP004308 Oryza sat
C 33	50.6	3.0 93419	2 AC121747	AC121747 Rattus no
C 34	50.6	3.0 159563	2 AC112892	AC112892 Rattus no
C 35	50.6	3.0 162473	2 AC127748	AC127748 Rattus no
C 36	50.6	3.0 166421	2 AC098016	AC098016 Rattus no
C 37	50.4	3.0 1285	14 AF202064	AF202064 Bovine he
C 38	50.4	3.0 125020	9 AF429315	AF429315 Homo sapi
C 39	50.4	3.0 134374	2 AC121465	AC121465 Rattus no
C 40	50.4	3.0 144404	8 AP003210	AP003210 Oryza sat
C 41	50.4	3.0 18764	8 AC087726	AC087726 Chlamydom
C 42	50.2	2.9 168532	2 AC121451	AC121451 Rattus no
C 43	50	2.9 2443	9 HUMRASR2	M30539 Human SK2
C 44	50	2.9 10197	1 AE008327	AE008327 Agrobacte
C 45	50	2.9 12006	1 AE009285	AE009285 Agrobacte

ALIGNMENTS

RESULT 1
AC024609/c 90341 bp DNA linear PLN 12-SEP-2000
LOCUS Arabidopsis thaliana chromosome I BAC F14P1 genomic sequence,
DEFINITION complete sequence.

ACCESSION AC024609.2 GI:7212002
VERSION
KEYWORDS
SOURCE
ORGANISM

Arabidopsis thaliana.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 90341)
REFERENCE
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,

Altafi,H., Nguyen,M., Lam,B., Southwick,A., Miranda,M., Brooks,S., Buehler,B., Chao,Q., Chin,C., Chio,J., Choi,K., Gonzalez,A., Homg,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M., Lenz,C., Liu,A., Liu,S., Mukharshy,N., Pham,P., Sakano,H., Shim,P., Toriumi,M., Vayberg,M., Yu,G., Becker,J., Theologis,A. and Davis,R.W.

Unpublished

2 (bases 1 to 90341)

REFERENCE
AUTHORS
JOURNAL

Altafi,H., Nguyen,M., Lam,B., Southwick,A., Beil,Q., Hansen,N.F., Chin,C., Chio,J., Choi,K., Dunn,P., Gonzalez,A., Homg,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharshy,N., Pham,P., Sakano,H., Schwartz,J., Shim,P., Thavert,A., Toriumi,M., Vayberg,M., Walker,M., Yu,G., Becker,J., Theologis,A. and Davis,R.W.

Direct Submission

Submitted (01-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

3 (bases 1 to 90341)

REFERENCE
AUTHORS
JOURNAL

Federpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Beil,Q., Buehler,B., Chin,C., Chio,J., Choi,K., Dunn,P., Gonzalez,A., Homg,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharshy,N., Pham,P., Sakano,H., Schwartz,J., Shim,P., Thavert,A., Toriumi,M., Vayberg,M., Walker,M., Yu,G., Becker,J., Theologis,A. and Davis,R.W.

Direct Submission

Submitted (09-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

4 (bases 1 to 90341)

REFERENCE
AUTHORS
JOURNAL

Federpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Becker,J., Theologis,A. and Davis,R.W.

Direct Submission

Submitted (11-AUG-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

5 (bases 1 to 90341)

REFERENCE
AUTHORS
JOURNAL

Federpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Becker,J., Theologis,A. and Davis,R.W.

Direct Submission

Submitted (12-SEP-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

On Mar 9, 2000 this sequence version replaced gi:7121532.
Bases 1-59,676 of IGF clone F14P1 overlap with IGF clone F6F9, gb|AC007797.

COMMENT

email for correspondence: arabes@sequence.stanford.edu
Genes with similarity to proteins in the databases are named 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The gene prediction programs used to predict genes include: Gtall (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genomic.stanford.edu/~cbr/GENSCAN.html>), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).

FEATURES
SOURCE

1..90341
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F14P1"

misc_feature
1..59676
/note="overlap with IGF clone F6F9, gb|AC007797, see GenBank record for BAC clone F6F9 for annotation in this region."
60210..61960
gene

gene
CDS
/gene="F14P1.1"
join(60210..60507,60608..60778,60878..60927,61025..61081,61188..61249,61387..61461,61599..61717,61807..61960)
/gene="F14P1.1"
/note="Unknown protein"
/codon_start=1
/protein_id="AA98407.1"
/db_xref="GI:9795589"
translation="MRSIQAPVPCSVPRPGLGVSALLVNCVSQKTRSLRQPMQNOTKNDSSQATVATLHLRLRYKSKICLPSHSDGSGTAEPNDENDYKSSVLEAVRYKSGDGMKNDGQRLCCHANNQGSNPNVAPSAVLKMGDTGLPLPITVLEMPSTVLAATVQIARPTMIVQVKWDKNGYEVRLVRYTTHVPAFALYLSKVGKSDCVSPDRSDALINAVRCVPIQVKYLAASDQKVIDGKLSQGTAPASGLFTELDNRNGQPCPTKEPDLVRNMQAVDEERYDRAGKLSLSHPYVLVILVINY"
61540..65756
/gene="F14P1.2"
join(61540..63596,63685..63951,64046..64123,64205..64306,64387..64436,64527..64626,64741..64833,64927..65040,65139..65253,65359..65756)
/gene="F14P1.2"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AA98408.1"
/db_xref="GI:9795590"
translation="MWTNMIQWRDRDPTLLEDFEPELDEVLARYYPQGYGVNKGKRPVYIEELGVDSKLMQVTLLERYHYKFEKTTIVKPPACCTAARHTDSSTTLDVQGLKPKTKTARDLILOKIDSDNPVETLHPMTINAGSGKTKGVSFLDPKYSKYHLYGNKYONKILMIDASOLDPPRGCTCDQCGCMRDKAPKQSEILKMGRCGCTPCRHGAELSDSQTSSSDPTYSLRSDSTASQSELEBASPKNTNNHVPCLTPVBSVLANGISPTVLSEBECVPMQKVDVAMQVQAPASGPDQYSSKIGSVRHVSWLTAFFISFPTLLASLALPQTEHQSLSSVYALCDERRIARESRPSPRSSTTVRYIISSVLSRLDLEKQIENLHRSKSEMPHEKEELNAAVYRDALAEALITTKYK"
70742..73284
/gene="F14P1.3"
join(70742..70816,70957..71382,72084..72173,72904..73284)
/gene="F14P1.3"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AA98406.1"
/db_xref="GI:9795588"
translation="MEVWRYLVHNNKNGETSYAKNSTAOSNIISLGRVMDALKLMSNBSISIGIADLCSSGNSLISINIVDIHNLDPDIDRVPETRLSINDI.PSNDPNYICASLPEDFYRVNNKXGLGRGGEGSEFYAIPSGFYRLPPRPSLPHASSSIHVLISOKTYSNRRRELISMTGTPSSSSIVHQGCIYIIEEKIDAPRAPIHASSSELMKVIKXGSPSIDRLERISIDWGGGSISSSDVLRKPEALSKRSVNTIRAVVEPMLPRTGENVMDLPERYAKIVGEYVYSSPRAIVILSVRAG"
74532..76417
/gene="F14P1.4"
join(74532..74694,74769..75090,75300..75449,75558..75735,75908..75987,76159..76245,76328..76417)
/gene="F14P1.4"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AA98400.1"
/db_xref="GI:9795582"
translation="MFQDRKSTAGVPGSGFPVIGETLQPMLSVNSGKGFYFVRSRRIVGSCRTSLGETHVPLSTGESARVLVANDSGMFAKRYKISGLVQDRSLCAPOHKKILSRILNLFKRSKTLAMVHPDELVAVLALGCMHGHTVALLDLOITTPAMCMGLVLSKEKEELSGMORVPGVCEAMLAFLPLPMPTRPHSVIARVWEMLEKIRRRRINSHNNHEDLIQOLAVNDPQOLIDEIENDITMIIAQDPTASLTMMVKIKGNQKVDLIIIEBOSQITKASNKPFLLEBLSNPYASKVKSLSHNASVIVMPPLVLIQDEMGILL"
78430
/gene="F14P1.5"
complement(77375..78430)
/gene="F14P1.5"
/note="Unknown protein"
/codon_start=1
/protein_id="AA98401.1"
/db_xref="GI:9795583"
translation="MEAEKASWYENLFOACKTYGVLSAVVAGYCCLPVIVQHLGPIWDFWFLVLRNLPFLSVVSIIFLIVIFSKIQTEKKEBDLYDQVVAVPP

```
STVTPDSYNSKPFEDVGYXNNSTYRENPAPFAQVAVESSEYSYDOEVEWMSKSR
RTSSSEKMTBEMVEKTERETRYTGTSGMSQSDGLSGSEFRRTIOFFYSQKRM
HITRASGDOLENGXTHRASGDVDPVHWEHAGAPMONDVPOLONGVPHWKMKG
SVSOLQSDVPOMONGSVLQLONDVPOMOTGSGYOLONGHEPONGIYOMKGRSDG
SGRRHRHGRSRKERO"
/gene="F14p1.6"
/contig="F14p1.6"
/notes="Similar to defensin, similar to defensin AMP1 (anti-
microbial protein)"
/codon_start=1
/protein_id="AAF98402.1"
/db_xref="GI:9795584"
/translation="MWAVEGRICRRSKTWGFCGNTGCDSDCKMERASHGACHQ
PPGAPCCYFNC"
complement(81552..83365)
/gene="F14p1.7"
complement(join(81552..81800,81920..82031,82136..82233,
82329..82436,82543..82581,82670..82741,82820..83068,
83225..83365))
/gene="F14p1.7"
/notes="Unknown protein"
/codon_start=1
/protein_id="AAF98405.1"
/db_xref="GI:9795587"
/translation="WVAEALPKCEAPLVLGLOPALIDNVAPYDWSLLDQIPGDRG
SLAVOXDEHMLKELDAHISVAPLKKMGAGSYTVAGLSVGVAITGIDAYDDE
OGOLFVSNMFPSCVSISSLRKKKSTOCCEVVDGSGRMTPLCSAVIKADLSK
EDFTGSKMLVLTAVINLOVQOAIIRAPKEGISLIDLASFEMVRSSSELRQLES
GNIDLCFANDEABELRQBQAGEBALEFRRRCRAVAVTLESKCIKRDKEVTH
ISAIGETVADTAGADLPASGFPLGLIKLSLECCVSGCSGYSVIRALGSEVTEP
NMOWMKQLQKGLPVPDIHN"
complement(85358..87513)
/gene="F14p1.8"
complement(join(85358..85648,86056..86207,86310..86435,
87001..87144,87399..87513))
/notes="Unknown protein"
/codon_start=1
/protein_id="AAF98404.1"
Query Match 16.6%; Score 282.6; DB 8; Length 90341;
Best Local Similarity 57.9%; Pred. No. 2,46-50;
Matches 702; Conservative 0; Mismatches 304; Indels 207; Gaps 3;
gene
CDS
469 797 AATGAGTTGCTGAAGACAAATGTCGAGAAAGAGTCTCAACGGAATATTCGTGAATC 856
46551 GATGAGTAGTGAAGATAGTATTCGCAAGAAATGTTCTCGTAGACCAAGTCCGAGATC 46492
857 CTTGGAGTAGAGTAGAGATCTCCGTTGTCGATTAATCAAC----- 900
46491 TCTTGGAGTAGAGTAGAGATCAATATTGTCATTAATAGTTAGTCATTTCCTTAC 46432
901 ----- 900
46431 ATCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 46372
901 -----GTTATCCAGAGAGAAAGGAGCAAG 924
46371 TTAATTTGCTTCTGACAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 46312
925 ACTTATTTCTTCAAGCATTTTATCAAGGCTTTGCACTCATCAACAG---AGAACTTA 981
46311 ATCTATTCCTCGAGGGGTTCCATGAAAGTCTTAAGTATCAAGAGACTTAAGAAACTTG 46252
982 AAGGCTTAGAATACAGTGTAGTGTGGAGAGATGTTAATGCTCAGACCAATTTG 1041
46251 AGTACCAACATGACATGCACTGAGTGTAGTAGAGATGAGAGGACAGACAGAAATTTG 46192
1042 AGACTCAGTTACGTGACTTTGTGTGAAGAACAGATTCAATGACCGTGTCAATTTTGTA 1101
46191 AGACAGAGCTACCAACTTTGTCAGAGATGAACTTCAAAAATTTGTCACCTTTGCA 46132
1102 AAGAGACATTTGGGAGAGGCGCCCTTACTTGGCAGCAATGATGTCTTGTAGAAATCTC 1161
46131 AAAAAAATGAAAGTAGAGACCATATTTAGACACTATTAAGCGTTCTGTGCCAAACTCC 46072
1162 AGGCGCTGGAGAGATCTTTGGAGATTAACAATTAAGTGAAGCAATGACCTTAAGTTCCAG 1221
46071 AAGCGAGAGAGAGATCTTTGGAGATTAACAATTAAGTGAAGCAATGACCTTTAAGCTTCG 46012
1222 TTTTGGGAGAGAGTGTGTGAAGGACCAAGAGATCGTCTGGAAGGCTCGACTGAGCTTC 1281
46011 TACTTGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 45952
1282 TGCATCTGCTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1341
45951 TCCACACACCTGTGAAGATGCGGTTTACCTTCTGCCAAAACTGTGTAACTGTGAGCA 45892
1342 GCCAGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1401
45891 CTACGTGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 45832
1402 TCATGAGACACCAATGCTGAGAGATCCGGCGGTGTGAAGATGCTCTGAAGAAAT 1461
45831 TCTTGAACATCATATGCTCATGAAATAGCTTGTGCTCAGAAAGTGTCAACATG 45772
1462 CACAGAGCACTC 1474
45771 CAAAAATTCATCTC 45759
gene
CDS
469 797 AATGAGTTGCTGAAGACAAATGTCGAGAAAGAGTCTCAACGGAATATTCGTGAATC 856
46551 GATGAGTAGTGAAGATAGTATTCGCAAGAAATGTTCTCGTAGACCAAGTCCGAGATC 46492
857 CTTGGAGTAGAGTAGAGATCTCCGTTGTCGATTAATCAAC----- 900
46491 TCTTGGAGTAGAGTAGAGATCAATATTGTCATTAATAGTTAGTCATTTCCTTAC 46432
901 ----- 900
46431 ATCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 46372
901 -----GTTATCCAGAGAGAAAGGAGCAAG 924
46371 TTAATTTGCTTCTGACAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 46312
925 ACTTATTTCTTCAAGCATTTTATCAAGGCTTTGCACTCATCAACAG---AGAACTTA 981
46311 ATCTATTCCTCGAGGGGTTCCATGAAAGTCTTAAGTATCAAGAGACTTAAGAAACTTG 46252
982 AAGGCTTAGAATACAGTGTAGTGTGGAGAGATGTTAATGCTCAGACCAATTTG 1041
46251 AGTACCAACATGACATGCACTGAGTGTAGTAGAGATGAGAGGACAGACAGAAATTTG 46192
1042 AGACTCAGTTACGTGACTTTGTGTGAAGAACAGATTCAATGACCGTGTCAATTTTGTA 1101
46191 AGACAGAGCTACCAACTTTGTCAGAGATGAACTTCAAAAATTTGTCACCTTTGCA 46132
1102 AAGAGACATTTGGGAGAGGCGCCCTTACTTGGCAGCAATGATGTCTTGTAGAAATCTC 1161
46131 AAAAAAATGAAAGTAGAGACCATATTTAGACACTATTAAGCGTTCTGTGCCAAACTCC 46072
1162 AGGCGCTGGAGAGATCTTTGGAGATTAACAATTAAGTGAAGCAATGACCTTAAGTTCCAG 1221
46071 AAGCGAGAGAGAGATCTTTGGAGATTAACAATTAAGTGAAGCAATGACCTTTAAGCTTCG 46012
1222 TTTTGGGAGAGAGTGTGTGAAGGACCAAGAGATCGTCTGGAAGGCTCGACTGAGCTTC 1281
46011 TACTTGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 45952
1282 TGCATCTGCTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1341
45951 TCCACACACCTGTGAAGATGCGGTTTACCTTCTGCCAAAACTGTGTAACTGTGAGCA 45892
1342 GCCAGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1401
45891 CTACGTGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 45832
1402 TCATGAGACACCAATGCTGAGAGATCCGGCGGTGTGAAGATGCTCTGAAGAAAT 1461
45831 TCTTGAACATCATATGCTCATGAAATAGCTTGTGCTCAGAAAGTGTCAACATG 45772
1462 CACAGAGCACTC 1474
45771 CAAAAATTCATCTC 45759
```

```

OY 797 AATGAGTTGCTGAAGACAAATGTCGAGAAAGAGTCTCAACGGAATATTCGTGAATC 856
DB 46551 GATGAGTAGTGAAGATAGTATTCGCAAGAAATGTTCTCGTAGACCAAGTCCGAGATC 46492
OY 857 CTTGGAGTAGAGTAGAGATCTCCGTTGTCGATTAATCAAC----- 900
DB 46491 TCTTGGAGTAGAGTAGAGATCAATATTGTCATTAATAGTTAGTCATTTCCTTAC 46432
OY 901 ----- 900
DB 46431 ATCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 46372
OY 901 -----GTTATCCAGAGAGAAAGGAGCAAG 924
DB 46371 TTAATTTGCTTCTGACAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 46312
OY 925 ACTTATTTCTTCAAGCATTTTATCAAGGCTTTGCACTCATCAACAG---AGAACTTA 981
DB 46311 ATCTATTCCTCGAGGGGTTCCATGAAAGTCTTAAGTATCAAGAGACTTAAGAAACTTG 46252
OY 982 AAGGCTTAGAATACAGTGTAGTGTGGAGAGATGTTAATGCTCAGACCAATTTG 1041
DB 46251 AGTACCAACATGACATGCACTGAGTGTAGTAGAGATGAGAGGACAGACAGAAATTTG 46192
OY 1042 AGACTCAGTTACGTGACTTTGTGTGAAGAACAGATTCAATGACCGTGTCAATTTTGTA 1101
DB 46191 AGACAGAGCTACCAACTTTGTCAGAGATGAACTTCAAAAATTTGTCACCTTTGCA 46132
OY 1102 AAGAGACATTTGGGAGAGGCGCCCTTACTTGGCAGCAATGATGTCTTGTAGAAATCTC 1161
DB 46131 AAAAAAATGAAAGTAGAGACCATATTTAGACACTATTAAGCGTTCTGTGCCAAACTCC 46072
OY 1162 AGGCGCTGGAGAGATCTTTGGAGATTAACAATTAAGTGAAGCAATGACCTTAAGTTCCAG 1221
DB 46071 AAGCGAGAGAGAGATCTTTGGAGATTAACAATTAAGTGAAGCAATGACCTTTAAGCTTCG 46012
OY 1222 TTTTGGGAGAGAGTGTGTGAAGGACCAAGAGATCGTCTGGAAGGCTCGACTGAGCTTC 1281
DB 46011 TACTTGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 45952
OY 1282 TGCATCTGCTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1341
DB 45951 TCCACACACCTGTGAAGATGCGGTTTACCTTCTGCCAAAACTGTGTAACTGTGAGCA 45892
OY 1342 GCCAGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1401
DB 45891 CTACGTGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 45832
OY 1402 TCATGAGACACCAATGCTGAGAGATCCGGCGGTGTGAAGATGCTCTGAAGAAAT 1461
DB 45831 TCTTGAACATCATATGCTCATGAAATAGCTTGTGCTCAGAAAGTGTCAACATG 45772
OY 1462 CACAGAGCACTC 1474
DB 45771 CAAAAATTCATCTC 45759

RESULT 2
AC007797/c 119942 bp DNA linear PLN 12-SEP-2000
LOCUS Arabidopsis thaliana chromosome I BAC F6P9 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC007797
VERSION AC007797.7 GI:7839909
KEYWORDS HMG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 119942)
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
```

REFERENCE	AUTHORS	TITLE	JOURNAL
3	(bases 1 to 119942)		
Altafi, H., Nguyen, M., Lam, B., Southwick, A., Miranda, M., Brooks, S., Buehler, E., Chao, Q., Chin, C., Chio, J., Choi, E., Gonzalez, A., Hong, B., Johnson-Hopson, C., Khan, S., Kim, C., Koo, T., Lee, J. M., Lenz, C., Liu, S., Mukharasy, N., Pham, P., Sakano, H., Shim, P., Tortum, M., Vayberg, M., Yu, G., Ecker, J., and Davis, R. W.	Unpublished		
2	(bases 1 to 119942)		
Federaplel, N. A., Palm, C. J., Conway, A. B., Conn, L., Hansen, N. F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Lucio, S., Schwartz, J., Shim, P., Tortum, M., Yokskala, V., Walker, M., Yu, G., Ecker, J., Theologos, A. and Davis, R. W.	Direct Submission		
Submitted (12-JUN-1999)	DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
4	(bases 1 to 119942)		
Federaplel, N. A., Palm, C. J., Conway, A. B., Conn, L., Hansen, N. F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bel, O., Buehler, E., Chin, C., Chio, J., Choi, E., Dunn, P., Gonzalez, A., Hong, B., Kim, C., Koo, T., Lee, J. M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharasy, N., Pham, P., Sakano, H., Schwartz, J., Shim, P., Thaver, A., Tortum, M., Vayberg, M., Walker, M., Yu, G., Ecker, J., Theologos, A. and Davis, R. W.	Direct Submission		
Submitted (16-MAY-2000)	DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT			
On May 16, 2000 this sequence version replaced gi:7547094, Bases 60,267-119,942 of BAC clone F6f9 overlap with bases 1-59,676 of IGF clone F14P1, AC024609 and bases 1-3,490 of BAC clone F6F9 overlap with bases 89,221-92,710 of BAC clone T20H2, AC022472.			
FEATURES			
source			
1..119942			
/organism="Arabidopsis thaliana"			
/db_xref="taxon:3702"			
/chromosome="1"			
/clone="F6F9"			
2448..4393			
/gene="F6F9.1"			
join(2448..2640,2712..2995,3068..3502,3589..3727,3808..3974,4064..4393)			
/gene="F6F9.2"			
/note="Similar to endo-beta-1,4-glucanase"			
/codon_start=1			
/protein_id="AA012562.1"			
/db_xref="GI:10086502"			
/translation="WNAKPSRCCCCSVFICVITLAIIVITFIRHNSHDDGSGKVNQYAAAKIMQFPIQSGKLENNIISWRGSGIKGGSASIDISKILVDADPHKRGPAFAATVLSIIIEYDQMSLMLDHPKADIKTTDFLINHSPNVLYIYQGDPTVAKCMDEPMTKRTLTIKIDTPTGTEVAETAAVAAASLVPEESDPTDSTLTLHAQKIDEPADNRRGSSVAIVPEVSYNSGYGDELAAASMLVATRDORYLDFVSGEENGFGNGSPSWMDNKLGGHILSLRFTFFKSGSGKIGCEPKRTAAVNGCGLPSPATVSSRRDGLIIVSENNAILGHVSNLAFATLIVSDYMLTSGYKELSCSDQSFKSDIDEPFASQDYMILKQNP EKKSGILVYSGKIPFVHRKASIPADATVCGKGGPKKINSDENPNVATGALVGGPFLNDITDRNNMSQNEPSTNSALVGLLSSLVTTSSSYEST"			
6059..7881			
/gene="F6F9.2"			
join(6059..6600,6686..7091,7183..7443,7547..7633,7747..7881)			
/gene="F6F9.2"			
/codon_start=1			
/product="sulfate adenylyltransferase"			

gene
CDS
/protein_id="AA012544.1"
/db_xref="GI:10086481"
/translation="MSIMIRSVSHVITLPQPNRSKPSGFMQISFLSSNNRPITL
VYKRLIMQSVSNMYKSSLIIPDGGELVELIYERIEIDVKKASSEMPVYKLTQDIL
EMWYHISCMASLKGPNREDEYLOSLHNSLRKNGITVNNSLPTVLADIDDTKEQI
GSSSEVALVEPQQDILIGLSKSVETIYNNKEIRLIRKWTSGPLVVEEITPSNNIL
IGDGLVEETIKNDGLDHRILSPKQLEEDNQQDAVFAQRLNPNVHGIALNNL
TKRRLLEGGKXNVNHLIIPAGGETKADVPIDJRMQHSQKILEQDGLPDKTTIYIIP
SPWHAIVQGVQWMAKARINAGANFYIVGRPPAMQHPTEKQDIYDPDQKXVLSMP
GLEKNTLIPRVAYDITEKKMAFPDPSRAKEFLISGTKNRYATGPNPDGPMG
SGMNVLKXYESIJSQSEKQCAAVSA"
/complement (8394, .9399)
/gene="F6F9, 3"
/length (8394)
/complement (join (8394, .8526, 8930, .9215, 9321, .9399))
/gene="F6F9, 3"
/codon_start=1
/product="vacuolar H⁺-pumping ATPase"
/protein_id="AA012542.1"
/db_xref="GI:1008482"
/translation="MASTFSGDETAAPFFGKAAALVPSGMAAGVGTAKSGVYAS
GMPREILVMKSGIPVMAVGLIYGLITAVISGTNPASAKYILPGVHALSSGLAC
GIAGLSGMAIGVGDAGVRAANQPKLFGVMITLILIPAEMLALYGLIILISLRG
OSRAE"
14164, .15678
/gene="F6F9, 4"
14164, .15678
/gene="F6F9, 4"
/note="Unknown Protein"
/codon_start=1
/protein_id="AA012543.1"
/db_xref="GI:10086483"
/translation="MEMQILHNDRLVVMYDRTNFGPSNLSLPGNCRDNPDAVSKIDC
TAHSIEVDANTTLEPLTVQSNVMSGSVVRPGVLYOTGGDIDRGLKTRTFSCNN
QODVWNNGLKKRWYASNHLPDQKQIVMAGGCGQFVPEFPKPTNPVVALDPLAE
THDQGGNNNTYFVNNMTDGNLPMFANRALLIDYKQVTVKTPPAIPEGCPNVVET
GSAVLDLPMLEADNVTYVLYVCGGAPKGSYNLAKRTFKVALDITCARLKNDAKPE
AYEKMPHAYKMDNIDIPDNEDVLINGSGFGTAMGRITVLAPOLVHPENPVGSF
ESLRPTIIRNHSKAILRDGRVLVGGSPHAFVNYGVGLPTEILSEASPVYLR
FSLNARLITSEPSQMIKTGNLKLRFVYAGVTTPKATVMFPTFTTHSPANQGV
LVNLVNFETKSGSMVEYVTRPSRANLAMPQYMI FVNVODI PSEGVWVLE"
complement (16924, .17682)
/gene="F6F9, 5"
17611, .17682)
/complement (join (16924, .17094, 17228, .17318, 17417, .17496,
17611, .17682))
/gene="F6F9, 5"
/codon_start=1
/product="putative histone H3"
/protein_id="AA012543.1"
/db_xref="GI:10086503"
/translation="MARTQIQTARKSTGKQPRKEATLAARKTRRPGYGVNAHFR
PGVALREIRKYSQSTDLIRCLPQRLVREILAQPFKIDLRFSQSHAVILQEAAEAYL
VGLIEDTNLCIHHKRVITMSKDIQLARRIRGERA"
18898, .22068
/gene="F6F9, 7"
join (18898, .19093, 19182, .19243, 19336, .19403, 19510, .19611
19701, .19824, 19937, .19988, 20276, .20393, 20887, .21016,
21098, .21220, 21301, .21354, 21480, .21661, 21734, .22068)
/gene="F6F9, 7"
/note="Unknown Protein"
/codon_start=1
/protein_id="AA012544.1"
/db_xref="GI:10086484"
/translation="MAEANNSEKKDDSSBEKGEELLFCGATAMDIIGKRKGAMEGNTL
VSPRLRLPQVNRIPVATGSCFPCVALLDVGRCYTMGNENKQGLGHGMDIORDP
VTSGLSKRIITKAAARNNITVYVSDGQSLGGMKKYGLGLGSAKNSFVSEVSTP
LPCVSDDEVYNNAGADPTVWLSSTBGASLIRAGIPQYQGLDGLHDFNNKASRYL
AYAGQRPKALASLAEETIYKVALGNTHTVNSITIKTKRLPLHSGGGQLYMNGKITN
NGMDMTRPMDVLSGMNLRMWDSSGWHFVADSSCSITGHAQVLEGGVGPNGQKS
APARKDILBSHWGTDLADRLQVTKARBEFFRLNKLKRYKTISSRGISSSYSL
HCLLCIQLEVDNGKSLSEASVEVEETLAPKQCAVTKRGASGRKRSVAASDSBOD
SDSDNSDEKXVQSDADSDYSEDGEANQKQSAIRGRGARGARGLSGRTSGKAP
VATVGRRRPRTKS"
24177, .27256
/gene

CDS	/gene="F6F9.8"	join(24177..24269,24579..25172,25331..25510,25592..25651,25766..27256)
	/gene="F6F9.8"	/note="Unknown Protein"
	/codon_start=1	/protein_id="AAG12559.1"
	/db_xref="GI:10064499"	/translation="MGSPASSSCRLISCSGDDTSDAPNSLTALENKSGDKRGMSPRKSQKORGLIVUSFTTPASRTRELSALKSPSPNNNSPEKXOOSFSYDEKQQLPVYTYAPVDEKTKTGQVIEKTELLSYEPDTHRENVPIVSEKTEHEEDLIGTELQGSADADAKIEEDVTEVENASKVPRESBDVITIRKSEPERKDELSYIVVIOAVRGFLARRELSKXVILQOAVRGHLVNSQMSGLFCVQAIYKQAMTRARHSTKDSRVATSDESEPNAAQKLENFPAHLMESTPKPIKICODTPRESSANNWLERMMSVPRPKTSKANLITTEBONLEETONVKSIPQADFPVNSDTEVNEKITMP3YBASKVEQUNVELSETEKMSQYDSEASAEVYDSIQOPLAANDLEEPYVYGQIKHSLKRVNSPFIAAOSKFEELISSTGSKAMTSLSKODVLAEGEKTDI06PDTVTIKHSLLEDVTPALISGSCGTELSVTSILDTLEKSDLEGADEPREVAKLIEDPFTKDQALILIDVXDATSMGTVEDPKKEVKNANDEVEISAHTHEPVI1STPD3KKRADEDSGPOAYALSBEGLTPMTITTSQATPASOASSVBARPKSEKSGSSOKRXYKSKITISPKOEITGATBEQEGKQSGRITSFGIDQABSSGCGNELPFPNQPOTSAAYKXOEHNSPSSPDIOERDVVASAKRHSILPGVTNGQVSPRIQSA5QAOQGTGKANLLOSFLCNHSSQP"
gene	28911..31062	
	/gene="F6F9.9"	join(28911..28985,29480..29677,29976..30100,30198..31062)
CDS	/gene="F6F9.9"	/note="Hypothetical Protein"
	/codon_start=1	/protein_id="AAG12545.1"
	/db_xref="GI:10064485"	/translation="MOENAPRTMSDVORVLAINKVFDMVLFISEDSPOVSQSSODHLOAKSEPLASHPEDDLPPGFGFSAFANESQIKLADILIMKCSVOILLDEMEVVAQDEKEVEAQNRELERLVEAFYPGASSIPENPSVPADVDESHDDQTVI1ILFVEDDILIMDSADFPTQSGVDGTEPSITIDBENTSTLPGPDIIMALLSALISNEKQSGMIDDLILKILSNKLENIUNVNRGSAGVSNSTSYSSTHEANVTTAATPISNGGFYAOPPIITHIPMAATYPPAPODOBPVGAAPARDAIYVNLILQOCHGRDEPTPVCHIGYRNLQPGGPNPBNVNSNNKQPRDSKRYKIMKACMFNSARCBRCRGNANCMYQDATTPYOPNLNNGNINTSDMNAKMRPDRD"
gene	complement(31753..35330)	/gene="F6F9.10"
Query Match	16.6%; Score 282.6; DB 8; Length 119942;	
Best Local Similarity	57.9%; Pred. No. 2.5e-50;	
Matches 702; Conservative	0; Mismatches 304; Indels 207; Gaps 3;	
QY	469 TGCAGTTTACCCAGCTAGAGACACAGAGGAGTTGATTCCTCTAAAGCTGATTTGG 528	
Db	107237 TTCAGGTGATATCAGCAAAAGTCACAGAAAGCTATAGATACAGCTCTCAAGTCTGATCTG 107178	
QY	529 TTATCTTAAACACGCTGCTGTCGCAAGTGGCTTGACCTGTTCTGAAAAGATCATGTC 588	
Db	107177 TTGTTTAAACCTGCTGCTGTCGCAAGTGGCTTGATGATGACAGTTCTCAAGACATGTC 1072118	
QY	589 CTAAAGTCTCTCCGAGAGATTTTGTGTGATCATGAATGCGTGGGCTTACTTTAAG 648	
Db	107117 CTAAAGTCTCTCTAAAGTACTCTGTGTGATTCATGAGATGAGAGGCTCACTAATTTCAAC 107058	
QY	649 TTGATATGTCAAACATCTTCCCTTGTTGCTGAGCGCATGATTTGTCATACAAAG 708	
Db	107057 CGGATTTAGTCAAGCACTTCGTTGTTGCGAGGTCATGATTTGATGACCGCAAG 106998	
QY	709 CTGAGTATTTGGAATGACGACCTAGAGATCGC----- 740	
Db	106997 CTGATACTGGAAGAACGACCTCATGATGCTTAGAGGTCTGTTTCTCTCTGAGAGTC 106938	
QY	741 ----- 740	
Db	106937 TTGCGATATATTAAGTTTACGTTATTTGTCATAAGCTTAAAGTGTTATTTATTGTT 106878	
QY	741 -----CTGAAATATACGAGCCACAAACTATATTTTCCACCTGGGGAAATAGTAAAGACT 796	
Db	106877 GTAACACAGGATTAATAATGCTTAACCTATGTCGTCGACCTAGAGAAATAGCAGAGACTT 106818	

QY	797 AATGAAGTTTGCTGAAGACATGTCGAAAGAGCTCTACGGGAAATATTTCTGTAATC 856	
Db	106817 GATGAAGTATAGCTGAAGATAGTTTTCGCCAAGAAATGTTCTCGGAGACAGATTCGAAATC 106758	
QY	857 CTTTGAAGTACGGAATGAGATCTCCCTGTTTGCAATATTAACA----- 900	
Db	106757 TCTTGAAGTACGAATGAGACACTATTTGGCATTTATATAGTTAGTCAATTTCTTAC 106698	
QY	901 ----- 900	
Db	106697 ATCTTCTCTCTCTAGTTTCTTTAAGATCATATCTGCAAAAGACTCAAAATTTTGG 106638	
QY	901 -----GTGATACAGAGAAAGGACAAG 924	
Db	106637 TTAATTTGCTTTGACAGTGCTCTCTTTATGACAGGCTGATCTCCAGAAAGGCGCAAG 106578	
QY	925 ACTTATTTCTTCAAGCATTTTATACAGGCTTTGCAGCTCATCCAAACAG---AGAACTTA 981	
Db	106577 ATCTATTTCTCCGAGCGTTCCATGAAAGCTTTAAAGTATCMAAGAGACTAAGAAACTTG 106518	
QY	982 AAGGCTTAATATACATGCTGTAGTTGAGGAAGTATGATTAATGCTCAGACCAATTTG 1041	
Db	106517 AGGTACCAACATGCTGCTAGTGTAGGAAGCGATATAGCGCACAGAGAAATTTG 106458	
QY	1042 AGACTGATTACGTGACTTTGTGTGAAGAACACGATTCATGACCGGTCTCATTTTGTGA 1101	
Db	106457 AGACAGAGCTACGCAACTTTGTCCMAGAGATGAACTTCMAAAATTTGTCACCTTTGCA 106398	
QY	1102 ACAAGACATTTGACAGTGGCGCCCTTACTTGACAGCAATGATGCTTGTTCAGAAATTCG 1161	
Db	106397 ACMAAACATGAAAGTAGACCACTATTTAGACGATATGACGTTCTGTGCCAAAATCCTCC 106338	
QY	1162 AGGCGCGTGAAGATCTTTGGAAGATTAACAATGAACATGACATTTAGATTGCCAG 1221	
Db	106337 AAGCAGAGAGAGATCTTTGGAAGATTAACAATGAACATGACCTTTAAGCTTTCTCG 106278	
QY	1222 TATTGGGACGCGTGTGAGAGGACCAAGAGATCGTCTGAGCGGCTGACGTGGCTTTC 1281	
Db	106277 TACTTGGGACCTGACAGAGAGGACCAATGAGATGTACTGAAACGAAACAACAGGCTCAT 106218	
QY	1282 TGCATCTGCTGGGAAGAGGGCGTGGCGCTTTTGCMAAGAACATGTCAGACTCGCAA 1341	
Db	106217 TGCACACACGCTGTAAGATGGCGTGTACCTCTCCCAAAAATCTGTGTAAAGTGGCGA 106158	
QY	1342 GCCACGCGGACAGAGGCTCTCATGAGGAGAAAGGGCTATGCGAGGTGAAGAAATGT 1401	
Db	106157 CTAACTGAAATGAGGAATATCAATGGGGAAGAAAGGATAGAGGGTTAAAGAGATGT 106098	
QY	1402 TCATGAGCACACATGCTGAGAGATCCGGCGGTGTTGAAGATCTCTGAGAAAT 1461	
Db	106097 TCTTGAACATCATATGCTCTCATGAAATGCTTGTGCTCAGAGAGTGTCAACATG 106038	
QY	1462 CACAGAGCACTC 1474	
Db	106037 CAABAATTCACCTC 106025	
RESULT 3		
AC023754/c	100685 bp	DNA linear
LOCUS	Arabidopsis thaliana chromosome I BAC Flib6 genomic sequence,	PLN 13-SEP-2000
DEFINITION	complete sequence.	
ACCESSION	AC023754	
VERSION	AC023754.3	GI:7212003
KEYWORDS	HTG.	
SOURCE	Arabidopsis thaliana.	
ORGANISM	Arabidopsis thaliana.	
REFERENCE	1 (bases 1 to 100685)	
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,	

JOURNAL AUTHORS	Altafi, H., Nguyen, M., Lam, B., Southwick, A., Miranda, M., Brooks, S., Buehler, E., Chao, Q., Chin, C., Chou, J., Choi, E., Gonzalez, A., Hong, B., Johnson-Hopson, C., Khan, S., Kim, C., Koo, T., Lee, J.M., Lenz, C., Liu, A., Liu, S., Mukharly, N., Pham, P., Sakano, H., Shin, P., Toriumi, M., Vayberg, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.
TITLE JOURNAL	Unpublished 2 (bases 1 to 100685) Submitted (17-FEB-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE AUTHORS	3 (bases 1 to 100685) Federpiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bei, Q., Buehler, E., Chin, C., Chou, J., Choi, E., Dunn, P., Gonzalez, A., Hong, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharly, N., Pham, P., Sakano, H., Schwartz, J., Shim, P., Thaveri, A., Toriumi, M., Vayberg, M., Walker, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.
TITLE JOURNAL	Direct Submission Submitted (09-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE AUTHORS	4 (bases 1 to 100685) Federpiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Becker, J., Theologis, A. and Davis, R.W.
TITLE JOURNAL	Direct Submission Submitted (13-SEP-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT	On Mar 9, 2000 this sequence version replaced gi:714318 Bases 1-32,040 of IGF clone F1B16 overlap with bases 97,508-129,547 of IGF clone F10A5, gb AC006434. e-mail for correspondence: arab@sequence.stanford.edu Genes with similarity to proteins in the databases are named 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://genomic.sanger.ac.uk/gf/gf.shtml), and NecPlantene (S.M. Høbsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NecPlantene.html).
FEATURES source	1..100685 Location/Qualifiers /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="F1B16"
misc_feature	1..32040 /note="overlap with bases 97,508-129,547 of IGF clone F10A5, gb AC006434. See Genbank record for BAC clone F10A5 for annotation in this region."
gene	/gene="F1B16.1" /complement(32986..33822) /note="Unknown protein; unknown protein"
CDS	/codon_start=1 /protein_id="F1B16.1" /db_xref="GI:10120444"
gene	/translation="MALPVSPPSTSLSPFSKRPFHSLSLSPVDRRRCKLN SSFRSSSSFSFEKHNAKSPKSDIVLELPPLPLVFPGLIPLOIFEFYIMAT LLOGLDRGVVSDVSGSAVGVGVHVEHRLVDRFFLVCGQRRPVRVVR KPLVGEVYLMLEDRSGEENLDSLANEVEVIMKEVIRLSRLNGAKAEVODLRROP PTPFSEFVGSTFEGAPREGOALLEEDPAARLKRBEETLRNTLVNTAASAVDVPS SS"
gene	43756..46560 /gene="F1B16.2" /join(43756..44347,44507..44562,44700..44966,45493..45755, 46200..46560) /gene="F1B16.2" /note="Similar to cytokinin oxidase"
CDS	/codon_start=1 /protein_id="F1B16.2" /db_xref="GI:10120443"
gene	/translation="MTSSFLITPFCIKLIANGLVGPEBELIGALIDVDGHFTVP SLASVSDPDMKSPBEFLAVLHPSABDVRLVATAGSATLFPVSARHGHSING QAAAGRNVAVEWNGHVGVTGTPKPLRPDEWVLDVWGLMDVYKTLLEHGLPKSWT DLYLVTVGTLNAGISQAFHGGPOISNLELDVVTGQFGIITRARISLEPAPORV WIRLVLSFVFTEDDEYLSWGHQLPFVDEGVFVVDGGLVNRSSPSPNPVKI SSVSSNGSVLYCLEITKXNDSDSEIYDQVLEIMKKNFTPTSPFTDLOVYFLDR YKAEELKLRKKNLMEVPHMLNLSLFPKRSISDPDKGFGLIGKSGPLIYPMNKD KNDERSAVTPDEEVPYVALRLSLDGGETOKLEYIKDNRILRFGOATVYVKO YPLHATQGEVVAHFGDKWDRFRLLKAEPPRIIATQGRIFONPSLTPSPSSSS AASW"
gene	complement(46848..48185) /gene="F1B16.3" complement(join(46848..46943,47040..47128,47474..47537, 47773..47875,47963..48066,48156..48185)) /gene="F1B16.3" /codon_start=1 /product="Putative ubiquitin-conjugating enzyme"
CDS	/protein_id="F1B16.3" /db_xref="GI:10120441"
gene	/translation="MSSGAPSRKTLATNRLOKEIVEMQNPPTGRHKVYTNLQ KMLIEVGAHGLVADVTLQVDFPEHPMESPVITLHAPLPHIIVSNGHICLDI LYDSWSPAMTVSSICITSLMSLSSSTEKORPTDNDRYKKNKGRSPKTRFWFHDK V"
gene	52291..53501 /gene="F1B16.4" /join(52291..52584,52917..53501) /gene="F1B16.4" /note="Hypothetical protein"
CDS	/codon_start=1 /protein_id="F1B16.4" /db_xref="GI:10120438"
gene	/translation="MEDPRVHBCSLRGTLDSRYAKAVACVLEVIDIGREYELC NNILINDLPGRRRPFPALSEIKSLSSGSPSLPENHEHLKITKLSLVQYBER PEOYCNOLFOVVISFEBIAGSGSKYTGALQAMTRFGLSEALISQNLVYRRRT ISHDVPKLISSGLSLSLPGMTSSLSIORGLVGGQHAKPRGIPETSVATIR AMLOHFLHPSVFLILKCNITSDBSCLKCMRNISLILAVKTYIRMKQKMWLKD PRRKTPYNTLILTF"
gene	complement(54491..56307) /gene="F1B16.5" complement(join(54491..55074,55197..55351,55498..55769, 55872..56014,56253..56307)) /gene="F1B16.5" /note="Hypothetical protein"
CDS	/codon_start=1 /protein_id="F1B16.5" /db_xref="GI:10120445"
gene	/translation="MGTDGTVSICINSYDSRGLPMLLELAFLRGVADVMTNQ KPLEDEVVYSLHEMLDRGVYISAKQKAVDLSLADILVNAVAGKMDLVLEK NVKVLPIKLMWITHEMRGHVFNADLVKHLPEVAAAMDLSHATGWKRVKRGGLIM PKTYVHLNLSKELEVAEDSVAKRLVLEHRESLGRNDDLFGIINSVARGGQDL FLAFLHESLERIKKKLOVPTMAVYVSDMSKOTKEETELRAGKMETLVNQTG KTLTVAPVYLAIDVLVONSQARGEGRITTEAMAFLVLTGTAAGCTMETLVNQTG LHSAKRGVPLAKNIVKLAIVDELRIRMGKNGYEVVEMFLHNSHRLASVLTKEV LQAKARALR"
gene	55918..61713 /gene="F1B16.6" /join(55918..60655,60734..61110,61188..61248,61315..61713) /gene="F1B16.6"

predicted gene.
Similar to maize transposon MudR mudra protein isolog
(AC003981) "

/codon_start=1
/protein_id="BAA92398.1"
/db_xref="GI:7228438"
/translation="WTSKGIIDRAERTFQSYUMLNGFRIDQEVYTVSVVSRAT
EGFMEIAMPDSTAMKRVYEMAFERSMPLVTFVVEKDIIVNSQTDVGGPIINAD
VVGPMQENQPREBOAMQMDADBERGVIQIIVDEKEDIDNEADASSDEGDVAM
ATWMANEDPSGLVIEBQDHPMEYKENEYIDGARYAHKDEMEKAVMEVAJDOEFV
VKSNTYVVEVRCMEKDCPWRHVAHKANNYWKVSYVTEHKCITGVGVKRYRNTSD
VASEMYSVWNIIGPEBKSIIRHIEKKEFYTSYVKAERKCKIEMEGYKRYSTADN
LPRLATIIAORNNYTDLHTFVSDDRTKSVLORAFSGICINAPHCPCPLCIDG
LPHDRAGMLRAIDYLONGDEKRLPAKPVYRSRQMRMGNAHYKQKKNHMEIP
KGLCANQOKERNEMLMDLDELTTKQDEQRPOVGDPEPIPLGALHDDPTMRR
SGSSINFTQMIENEPTKWSLLPTDSDRYGIMTNLAENVMTMRGVRLPLVAIV
EPLHGTQAVFRDVRKKGISMANNNIVGFEQVMRLPHLGPANTLPEYTVGSKAM
LLOPTADLHPDADNNSVRSLEAYLLMFGWMTSTGHADVRLVHARSIVAO
PODVQMSWGSVAIVATYALCEACTKDAGAIJAGCPMLQDMAERFAIGRPVDS
APYVGRSAQMPEDGPTMGTYMCRGRGRYAVOVRGYPDVFEEDRLQPDYIMEPY
TEBAVAAAPALGLSLCTRDQAYMLTLIPWFDIYVPHPCQORVWROGLRQVPSNY
QPTVLPADHSLTRRGQALAGLAPVQYVVDVTLATEVINELEFTEENYKDYLM
YLPRTARVETPTDAPPRHAYVTHAIFTHRDYFVGADARDISADITAVOYRLRM
GHLTDVEQRTVTRDQEKRAVMRVFSCRAVDIVPPGPDVCLRAPIASBQCLQ
HRTRTYPCNCTEIVWDMENPDQDTRAYRSETRDYLARKDHECRILAAALH
VNEGLPTWRREPCRCGRCHVIRSVRHRTGRCQVCPNIUDDPFIHAMWEYER
RVEIASLDMRNNPLGLPKWSEPRKCGCDRWQVATSPAQTFRRRYPVCPVNDDEV
CLTIYKRYDLHKITQTG"

/note="3' LTR"
/join(23418, 23748, 25558, 25589, 25715, 26480, 26638, 26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92399.1"
/db_xref="GI:7228439"
/translation="MGANCCIAAKERPOPCVTFIEVSAPFRVRSQSPMSFRMDNRTHI
EDIMEMPALFENHSSGIRPETKSGSIAPTDGSPNGSPDMFNKLKHSQDKRSGS
KIARSDLRGRTTSSNSPEAKLSKSLDITVSAQSKMSISVPEIPATIRADPSS
RGHSLPTADSMRKARSPQYLOVSDKIPSLSLMEGASPEGRSSMLSVCSN
DLSAVGSHGESSDQSGMRTPEBMAVSQSRREMYDEBLGVSYSKTRSNASNPVTH
SPDOVCYLCIKLKERSTWNAQBLAVVALGCHYHADCDJSTVLEADKXYPVCP
CTHGRKCTYKLFGLSKSTKTKIPKNVIVDNLGSSSKHQKEKIPLEELYSQARSM
RELKSSRIGANDMTLNDLRTEGFS"

/complement(join(27953, 28624, 29565, 29918, 30369, 30646, 31026, 31191))
/note="ESTs C96615(C10106), C2636(C12127), D21959(C10106)
correspond to a region of the predicted gene.
Similar to NAM (A021889)"

/codon_start=1
/protein_id="BAA92400.1"
/db_xref="GI:7228440"
/translation="MESLRDVLVLPFGFGFPHKDTELISHYIKKKIHQKKEVEIIPV
DIYKHEPMDLPACDVPTQDNKMFPAARDKYPNGSSNRATYAGWKTGSDRAIK
MGKQITGKTKTLVFEHGRPTGRTEMIMEEYIIDRECOAPCKMAVYLCRITKN
DWIPNGNELDNDSDPHEPYDAPBSVISTEQLAPAEVYGEAAPTGAALPGYTS
ATTANISPSDINDLDMELFDPFPDPOSLSASDLSPDEQVSSNVAALAKAVX
ODYSPNENVDDETYLLDEPVDYLIHPGTDPMNLQNPLOFTIQTAVDWSGQKS
ELMSPQANAPPSQSNAAINDGIRIRHSKTPETISYPOKGTAKMRVGINKMTSS
SESTINQITKENSRLVHEQKQADHVASIKTRSDAKPSTELSSNRGLFGRINAFAG
CSDARMMILVAGFALGVAVALHIGQRLGLSORDOHT"

/complement(38747, 38998)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92401.1"
/db_xref="GI:7228441"
/translation="MMGHPYLRAADTAPDWMNAVGEPEYRVHRHGTPEVYIVLCRPD
DRGCRASLMAAMMGLGVFEGCCYDDEPHGPGSTSI"

/complement(39315, 39726)
/note="3' LTR"
/complement(join(40209, 40396, 40917, 43545))
/note="EST C28952(C62945) corresponds to a region of the
predicted gene.

Similar to maize transposon MudR mudra protein isolog.
(AC003981) "

/codon_start=1
/protein_id="BAA92402.1"
/db_xref="GI:7228442"
/translation="MSKVTFOIVHGBNIRPGDVSIDPSVMTSGKIDRPAERTFO
SIYSWILRGPRIDQEVYTVSVSVVSRATGTYFMEIAMPDSTAMKRVYEMAFERSM
LIVFVSVEKDINVMQTDVEGPIINAGDIVGSPMQENQPREBOAMQMDADBERG
IIVDEMEBQSDNEVDNDASSDEGQVMTADNADSGEGLVITSEGHVMEYKENEV
TEGARVAKQMEKAVKHAUSIDREPRVYKSTNYTVYERCKEDCWRAHAKGMN
DYKWSYVTEHKYLDGCRKHNITSAEMASNTSVVNGIGEPSLIRHIEKFK
YITTSAAKRAKQKIIEMKYFEASDYNLPLRLAIQNNNTYDLHPTISVDRT
KSVLQRAVFSLGACINAFHCRPVLCIDGTFMGXKQIILTAIGCNGNOVLPAALA
PVESENESWYFLEERVHIAVVRPNVCLIHRRHAGMLRAIDYLONGMEKSLPAK
PDVRKQCMHMGANFYKQFNKLMELFKLCAQOEKFNELMDLDELTTKQDE
QRRRLVEGDEPEPIPLGALHDDPTMRREGSSINFTQMIENEPTKWSLLPTDGS
RGIMTNTNLAENVMTMRGVRLPLVAIVFELHGTQAVPPDRYKRTGSMADNIVP
GNVYTKYMEKIKKARRRVVAAQGYHREYKICVDNRRCITYKQAVQECYLKADG
CTCSQMPKTHHPCSHVYLAAGCGCLPPRYTVSNFKKAIPTHTMBEIVGGIISG
YITTSADVFTIPDPSKLRVKGKQRTNRINMDSEAGKRTKSCDRLGHTYKAC
PNAEVSFGADAPSGQASDVLFLHMYLQMLTTTHRRCTVSDPVARTVHNSTHE
KQULELSLTFPGNAYLHKVNAHA"

/complement(43999, 44385)
/note="5' LTR"
/join(44572, 44723, 44846, 44938, 46366, 46378)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92403.1"
/db_xref="GI:7228443"
/translation="MALPRGSPFKPMMAAASAPQPRGHMLVLTGFGVGRV9Q
RLQAGKRVSGTCTSPARKTELEMGWDAISFDATSSRCW"

/complement(join(48119, 48257, 48429, 48703, 49975, 50139))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92404.1"
/db_xref="GI:7228444"
/translation="MALYNDQSDSDYVNAACARAAPELVGAPMGFVQLARIGEC
EPPRRFRKKAITFCALSLFSSNRTITVLCYLRKNRPALIVYKESRTRCRRIT
GSLNARGDELICLARLASALDVRPKILKTSDFVLWLRKICVINPVHISRTH
GSSWYSLTDPVAVYRSTRELARGIVIDCTIGQ"

/complement(join(51918, 52637, 52684, 53346))
/note="Similar to Arabidopsis thaliana chromosome 1 YAC
YUPH1.2R sequence, C2-HC type zinc finger protein
C.e-WY1. (AC002986)"

/codon_start=1
/protein_id="BAA92405.1"
/db_xref="GI:7228445"
/translation="WASCIINGELHVRVHRRLYKASDESIRPHVLANSNLDLIPRTI
QVSNFCTYPRSTGDPFDVVAAPAGLPELHFFPLAQRIVSNPSCGLPEIHCNQ
GAEIVGADVALASLDYGVGASVGKILLPYAGVVALSVQVVSFACGGFTVANGTNA
VVDGSLGMLVSAWSEIARSGTLAAGARPNDRSVFRPSPSYGASLDEAFPLUG
ARQADRIAREQARTRVOAVAYLMKRLAAVYSGSDACRWMAVYDERRLLTSS
PELRAMRSYVGNVNTTPPAVATMERICQPLAAYASMAADAIARAAYEHPQELVW
VEHKGKORYIDTASVGLSPVSTTRAFSEVDITPGHAAKMLPTSSSSRLCT
GFVQIAARGGGDSWILSLMLPRLAALDESDELRIRKPVTAEYIGLRPASGSSA
KRAGMTSTXI"

/complement(57570, 58009)
/note="3' LTR"
/complement(join(58018, 58899, 58980, 62096))
/note="Similar to Arabidopsis thaliana chromosome II BAC
F903 genomic sequence, putative retroelement pol
polyprotein. (AC006248)"

/codon_start=1
/protein_id="BAA92406.1"
/db_xref="GI:7228446"
/translation="WRVNLQAGLMTAIDPGYAFREDMAALSAILOAVPREMLRGIA
KHDTARALADAIKTRVQVLRVVEAEOGFQRESRFRERTPEEFAMTLTVAVD
IRMGQVMEDEHVNKLLRVVPKKYPAVAILBOLDLVKTMLEIRVLRISTVYS
DEESGSGGLYLTBEQNAQVROKBOGSGNKGKGGAGQNGRPGGSPGXSA
ATGANSNRISRVKCFNCDFGHYAKCRKRRRORREANLVQAAEBEPTILMAHYVG

Query Match 13.2%; Score 224.8; DB 8; Length 146081;
Beet Local Similarity 70.2%; Pred. No. 8e-38;

Matches	369;	Conservative	0;	Mismatches	57;	Indels	100;	Gaps	1;
Qy	744	AAATATCAGATCCCAAACTTATGTTGTTCACTGGGGAATATGATAAACAATATGAA	803						
Db	123912	AGAAATCAAGATCCCAAACTTATGTTGTTCACTCGGGAATATGATAGAGGATTAATGAA	123853						
Qy	804	GTTCCTGGAAGCAATAGTCGCAAGAAAGCCTACGGGAAATATTCGATGATCCCTGGA	863						
Db	123852	GTTCCTGGAAGCAATAGTTGTCGAAGAAGATCCTACGGGAGCATATTCGTGATGTTCTTGA	123793						
Qy	864	GTACGAGTCGAGATCTCTCGTTTGGCAATATTAACA-----	900						
Db	123792	GTACGAGTGAAGATCTGTGTTTGGCAATATTAACAAGTATGCGCAATTGACGTTCTT	123733						
Qy	901	-----	900						
Db	123732	ATCCCTGTGTTATTCACATATATATATATATGATTATTAACAAATGATTTCTCTGA	123673						
Qy	901	-----GTATATCAGAGAGAAAGGACACAGACTTATTTCTTCAGACTT	943						
Db	123672	TTGATGTTAAATAACAGGTGTTTTCACATGGGAAAAAGGCGAGACTTATTTCTTCAGCGTT	123613						
Qy	944	TTATCAGCGCTTTCAGCGTCTCCACACGAGAAACTTAAAGCGCTGAATATCATGCTGT	1003						
Db	123612	TTATCAGGGGTTCAGGCTCATTCGACAGAAAGATTAAAGTGCCCATGATGATGCTGT	122553						
Qy	1004	AGTTTGGGAATGATGTTATATCTCAGACCAATTTGAGATCAGTTACGTGACTTTGT	1063						
Db	123552	AGTTTGGGAATGTGCAATAATCTCAGACCAATTTGGAGACACAAATTAGGTATTTTGC	123493						
Qy	1064	GGTGAGAAACAGATCATGACCGTTCATTTTGTGTAACAAGACATTTGCACTGGCAGTGCC	1123						
Db	123492	GGTGAGAAATGGGATTCAGAGCCGGTTTCACTTTGTAACAAGACATTTAGCAATGCTCTC	123433						
Qy	1124	TTACTTGGCAGCAATTGATGTGCTTGTTCAGAAATTCACGGCCGT	1169						
Db	123432	TTATTTGGCTGCACTGATGATGCTGTGTTCAGAACTTCACGGAGAT	123387						
RESULT 5									
AY091763		2557 bp	mRNA	linear	PLN 13-APR-2002				
LOCUS	Arabidopsis thaliana	At3g15940/MNC7	mRNA, complete cds.						
DEFINITION	Arabidopsis thaliana	At3g15940/MNC7	mRNA, complete cds.						
ACCESSION	AY091763								
VERSION	AY091763.1	GI:20147190							
KEYWORDS	FLI_CDNA								
SOURCE	Arabidopsis thaliana								
ORGANISM	Arabidopsis thaliana								
REFERENCE									
1	Shim, P., Chen, H., Cheuk, R., Kim, C. J., Meyers, M. C., Banh, J., Bower, L., Carninci, P., Chang, E., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jones, T., Kamuya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C. C., Toriumi, M., Wu, H. C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.								
2	Arabidopsis cDNA clones								
TITLE	Arabidopsis cDNA clones								
JOURNAL	Unpublished								
REFERENCE	2 (bases 1 to 2557)								
AUTHORS	Shim, P., Chen, H., Cheuk, R., Kim, C. J., Meyers, M. C., Banh, J., Bower, L., Carninci, P., Chang, E., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jones, T., Kamuya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C. C., Toriumi, M., Wu, H. C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.								
TITLE	Direct Submission								
JOURNAL	Submitted (26-MAR-2002)								

COMMENT	FEATURES	source
(SIGMAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA		
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RALF cDNAs (RALF cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,M., Hayashitaki,Y. and Shimozaki,K.		
The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RALF cDNAs: Shimn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Bahr,J., Bowser,D., Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Torigami,M., Wu,H.C., Yamada,K., Yamanura,Y., Yu,G., Yu,S., Davis,R.W., Theologis,A., and Ecker,J.R.		
Shimn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shimozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.		
Location/Qualifiers		
1..2557		
/organism="Arabidopsis thaliana"		
/db_xref="taxon:3702"		
/chromosome="3"		
/clone="RAF105-16-B09 (R21838) "		
/note="ecotype: Columbia"		
353..2446		
/note="glycosyl transferase"		
/codon_start=1		
/product="AT3G15940/MYC8.7"		
/protein_id="AA1M1011.1"		
/db_xref="GI:201471911"		
/translation="MEDIRLSPKIKSGSKSLSGSTPGSPGSLTRKHISGRTPRRRGGSGGAVVDFRSRLRLTWLILTLWTYLGVYQSRMAHDDNKVLFELFGKLRDELVLA		
VEQNRKADVDESSHAVVDHTNYVHALGVNRMHVIADKEDSTSRSVSPRRTRKKA		
SRSRSTKISTQKRWKMETKELDEQDELPNINIVYGLKPEGPSLEIRILEMSPQKRS		
RSQCDKSPKRLVWSRFPVLLFHELSMTQAPISMELASLLSCGATVAVVYVSR		
GLGLQELTRRIKVVDEKGLSPKTNAMKADVLAGSAVQASIDQWQVGGSGQIL		
WVWNNRPPRAKAPVLVDKVLILFELSDNSQMTLWEEDHDAKLSQDPVLPVLSL		
DELAFAVGASSSLNPTLTDETKRKRKLRESPRETRSDTKMYSLSLNGKGG		
QULILSVALLERQTOBOYAKNSQSKILKINLGIKRSKLSLSKARLVLRGSRTRKKA		
TSPVADHPSVLSATGRKLLSGNVTQKQDLKLLGVSQSKSVARRLVVEMLSPLSL		
NGNLSINSVLTPATTRVSLVSYSAADVNTYLGVEGTFGRVITENAAVGLPVGLTDG		
GKTEIVHNVTQLHPVGRAGNKVLAQNLIFLLRNPSRTLQSGRSEIVERKVKQKH		
MYKRFVDVLVTKMRP"		
BASE COUNT	730 a 479 c 625 g 723 t	
ORIGIN		
Query Match	5.3%; Score 89.4; DB 8; Length 2557;	
Best Local Similarity	54.8%; Pred. No. 9e-09;	
Matches 177; Conservative	0; Mismatch 146; Indels 0; Gaps 0;	
1116 GTGACCCCTTACTGGACGAATTGAATGTCTTGTTCAGAAATTCACAGGCCTGAGAA	1175	
2099 GTTGCTCATTTATCTCCGACAGATGTCCTAGTAAACAACCTCCAGGAGTTGGTGA	2158	
1176 TGCTTTGAAGATAACATTGAAAGCAATGGCATTCAGATTGCCAGATATGGCAGCGCT	1235	
2159 ACAATTGGGAGAGTGCATTCGAAGCAATGGCTTATGGCTCTTCGGTGTGGAAACGAT	2218	
1236 GCTGAGGAGCAACGAGATCGTCTCTGACGCGTGAAGCTGGCTTTCGATCCGTGCGG	1295	
2219 GCTGAGGAGCAAAAGAGATAGTAGGACACAAATGTTACAGGCGCTACTTCATCTGTGGG	2278	
1296 AAGGAGGGGTGGGCGCTCTTGCAAAAGAACATGTCAGATCCGCAAGCCAGCGAGAG	1355	
2279 AGGCGAGTAAACAAGTTTGGCACAGATCTTTGTTCTTTTGAAGAACCACTACA	2338	
1356 AGGATCTCCATGGGGGAAAAGGCTATGACAGGTGAAGAAATGTTATGAGACACAC	1415	
2339 AGGCTCAACATGATGATGACCAAGACGTAAGATCGTTGGAAGATGTAATAGACAGAC	2398	

NRKREKESLIGHRVLTSPISASTSESSASTSEVETTLRAKINELRGUATBELVTRKDF
GOLITIASPEELASPISTYISRLAKRVNIVEGELASERHIAQVRAMLEKATK
OTAVDHPATAKKAOEDALFVETIKKLALKEAKSVGKEFTVKELSHLIA
DASEAYEHGAVKARKOEDAOAEFEKATQSAELIWWKFLSLSL"
complement (14967. .17884)
/gene="F19K6.15"
/note="Identical to GI:6651430 from [Arabidopsis
thaliana]"
complement (join(<14967. .15347,15431. .15542,15637. .15739,
15870. .15904,16005. .16219,16303. .16412,16496. .16848,
16941. .17018,17169. .17244,17341. .17396,17478. .17547,
17650. .>17884))
/gene="F19K6.15"
complement (join(15143. .15347,15431. .15542,15637. .15739,
15870. .15904,16005. .16219,16303. .16412,16496. .16848,
16941. .17018,17169. .17244,17341. .17396,17478. .17547,
17650. .17823))
/gene="F19K6.15"
/codon_start=1
/product="beta-glucosidase, putative; 17823-15143"
/protein_id="FAG51546.1"
/db_xref="GI:12323126"
/translation="WVREKVALVGLALVLTVGAPTRKAGPYCAGLPDKESTIN
PEGRITGTATAAFOVGAVNEGCEGSMWDTFKPFRCEENADVAVDFTFRKED
IOLMKDINTAPRLSIAMPRIFEPGRMSKGI SKVGVQFYHDLIDELKNNIPLVTVF
HMDTPDLDEYGGFISGRIVODETEYANFTFHEGKWKMTIDFENPWFVRAGYDN
GKAPAGCSPIYIGYGOCDGSGEYAYOVSHNLLSHAYAVAFNCCOCAGKIDN
IASHPAMEPODLDEHYSIERVLDFTLGNHLPTTGYSVQSKDQVGRJLPKTEEA
EKLIKSTDYGMNYYTSVPAKEI SPDPKSPMTDLSLVMDKSDVGRKSGKSPFN
SKLIYVSKGI RYLKRYIDNQGDEPVI IANNGGDEKKNINDREGDNRKRYIIO
RHLSMDATCDKRVNVTGYFVWSLMDNPFWMQGYKARFGIYIIDPNNLTTHQVSG
KWSSEFLKQFPTSKREEL"
20839. .21105
/gene="F19K6.2"
/note="Predicted by Genemark.hmm"
join(<20839. .21024,21076. .>21105)
/gene="F19K6.2"
join(20839. .21024,21076. .21105)
/gene="F19K6.2"
/codon_start=1
/product="hypothetical protein; 20839-21105"
/protein_id="FAG51548.1"
/db_xref="GI:12323128"
/translation="WVDIYCAVAKGSGAGRNLLLIANASSDDQFWPFLQGLRYGFK
LFATIPDRLDECLTAGIEPTGSD"
complement (21772. .24076)
/gene="F19K6.4"
complement (join(<21772. .23135,23968. .24076))
/gene="F19K6.4"
complement (21772. .23094)
/gene="F19K6.4"
/codon_start=1
/product="unknown protein; 23094-21772"
/protein_id="FAG51549.1"
/db_xref="GI:12323129"
/translation="WGDSENVQPSKRRGALKOLSPDNGLDDNDGAELESTPKKA
SDEVLASRIYRKRPSPAPASNPFAIQIVTTTPASTPRTGNAPLASKLAP
AAVAVEDNOKASDI EBGDEYDSKVVDADVAGEBTEKXKDNKDCGASADVVAITE
VAQVSCDTNVCNNAVEGTDQDPLEKDSGODAKRKKEGNGIEEDNNGNAGFS
SFOQSSNRKNAFTGLASTASGSGSPGLVSGDGSCTSLGFLPSPSSSIIPAT
GSSIIKSGSGPPEQVSTETGEBENKVAFSADSI MFEYIDGCMKRGKGLKVN
SSNDKARLVRAKNGYRIILNALSYPKMLKMMKKGITFACVNSVSEGEGLSTFA

Db 2331 GGTGAACATTGTTGTAAGTACATGACAGACATGCGTTATGACTTCGAGTGGGA 2272
Qy 1230 ACGGCTCTGAGGAGACACGAGATGTCCTGACGCGCTTCGATCTTGCATCT 1289
Db 2271 ACAGTCAAGAGAACCAAGAGATGTTCAACACACATGATGCTCTACTTACCTCA 2212
Qy 1290 GCTGAGAGAGGCGGTGGCGCCCTTGCAAAAAGAACATGCGACATCGGACGCCGCC 1349
Db 2211 ATGGAGATCCCGTACAAAGAAATTAGCTCAATCTCTGTATCTACTTGAAGATCA 2152
Qy 1350 GAGCAGAGGCTCTCCATGGGGGAAAGCGCTATGCGACGGTGAAGAAATGTCATGAG 1409
Db 2151 GACGAGGCTTACGATGGGGAGTGAAGGACCGCAAAATGTTGAGAAAATGTACATGAG 2092
Qy 1410 CACCAATGGCTGAGAGATGCGCGCGTGTGGAAGATGTCCTGAG 1456
Db 2091 CACCAATGTACAGGAGATGTCGATGTTCTGTCATCATGATGAG 2045
RESULT 7
F6D8/c 117737 bp DNA linear PLN 18-SEP-1999
LOCUS Arabidopsis thaliana chromosome 1 BAC F6D8 sequence, complete
DEFINITION
ACCESSION AC008016.2 GI:5881519
VERSION AC008016.2
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 117737)
REFERENCE
AUTHORS Vysotskaya,V.S., Schwartz,J.R., Yu,G., Toriumi,M., Lenz,C., Liu,S.,
Lee,J.M., Li,J., Gonzalez,A., Liu,A., Liu,K., Mukharsky,N.,
Sakano,H., Vaynsberg,M., Chin,C., Choi,E., Chlou,V., Altati,H.,
Araujo,R., Brooks,S., Buehler,E., Chao,Q., Conn,L., Conway,A.B.,
Dunn,P., Hansen,N., Hwang,B., Hultzer,L., Khan,S., Kim,C., Palm,C.,
Rowley,D., Shim,P., Walker,M., Davis,R.W., Ecker,J.R.,
Federley,N.A. and Theologis,A.
Arabidopsis thaliana chromosome 1 BAC F6D8 sequence
Unpublished
2 (bases 1 to 117737)
REFERENCE
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-1999) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 117737)
REFERENCE
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1999) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 117737)
REFERENCE
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-1999) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
On Sep 14, 1999 this sequence version replaced gi:5441914.
The sequence of BAC F6D8 from Arabidopsis thaliana chromosome 1.
FEATURES
source
1. .117737
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F6D8"
/complement(1. .400)
/gene="F6D8.1"
complement (join(1. .20,293. .400))
/note="Similar to gb|Y12575 histone H2A.F/Z from
Arabidopsis thaliana. (This gene is cut off.) ESTs
gb|T14101, gb|T42178, gb|T43143, gb|Z26465, gb|Z28692 and


```

gb|743615 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55592.1"
/db_xref="GI:5903033"
/translation="MSGKAKGLIMKPGSGDKDKKKPITRSGRAGLQDPCELNS"
complement(831..2943)
/gene="F6D8.2"
complement(join(831..1072,1299..1438,1534..1671,
1782..1957,2322..2403,2509..2644,2826..2943))
/gene="F6D8.2"
/note="Similar to gb|A010025 unfr-interacting protein from
Homo sapiens and contains 3 Pf|00400 WD40 domains. EST.
gb|745021 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55591.1"
/db_xref="GI:5903032"
/translation="MEKKKVTIPQVCHGRSPVVDLPYSPITPDGFLISAKDSHPM
LRNGETGDMITGTEBGHKGAVWSSGLDNNALRAASADPSAKIMDALTDYLSPEBK
HIVACAFSBDTKSLTNGPEKILRPDMRLDAPPEVDRKSGSIRITVWLHSDQIT
LSQCTDGGVRLMDVRSKTIYQITETKSPVTAESQDGRITTAAGSTVFWPANH
GLYKSDMPCNTESASLEPFSGEFVAGSEDMWRVDPYTGEGICQNGHNGPVCV
RFTPTGASTAGSEEDGITRLWQTPANPENETSSRRVKSVDVEKSLIEGPHNKEG
KTAKEPSDT"
complement(449..4802)
/gene="F6D8.3"
complement(449..4802)
/gene="F6D8.3"
/codon_start=1
/evidence=not_experimental
/product="F6D8.3"
/protein_id="AAD55593.1"
/db_xref="GI:5903034"
/translation="MALITCSALPTIRASGSSGLNDPNRRKSAAMWAPLPGSPD
PDYLTNIESSCTVNPDKTIDISGSGQKFRGCTEKKAKQLRKTKAEASTPMDVWYSA
IASRLADITGRVED"
<5794..7183
/gene="F6D8.4"
join(<5794..5844,6737..6803,6864..7006,7136..7183)
/gene="F6D8.4"
/note="Similar to gb|D85381 cytochrome c oxidase subunit
Vb precursor from Oryza sativa. ESTs gb|R30504 and
gb|AA598195 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55594.1"
/db_xref="GI:5903035"
/translation="GRRLDIDHPSPFGTESPAVQSYFDPKRNIGCRGGEAGAC
IHLPIICVLISPTAEDGHDVWFMLDKGKSFPCVCSQYFEHGPDPGHDDDDHHH"
12444..14448
/gene="F6D8.5"
join(12444..12474,12682..12793,13387..13423,13523..13592,
13689..13793,13908..14042,14351..14448)
/gene="F6D8.5"
/codon_start=1
/evidence=not_experimental
/product="F6D8.5"
/protein_id="AAD55595.1"
/db_xref="GI:5903036"
/translation="MSYSHQSMGSGSRNARGERYVVRPKHQATLWMLHGICD
NGSSSSQMLDILHPNKKWICTAPSRPVTSIGPCTAPMDPVGEISPDGDDLEID
ASASHANLISSEPADIGSPMGAAISVATCYALGKYGNGAHYPIINQAVVGLSG
WLPADVDVYRFEKSKSQSGAGRLMFRPEYEG"
15655..17201
/gene="F6D8.6"
join(15655..15664,15795..15849,16134..16170,16281..16350,
16491..16595,16702..16825,17117..17201)
/gene="F6D8.6"
/note="Similar to F6D8.5."
/codon_start=1
/evidence=not_experimental

gene
CDS
13926..20924
/gene="F6D8.7"
join(13926..19355,19466..19583,19895..19931,20042..20111,
20228..20332,20406..20531,20797..20924)
/gene="F6D8.7"
/note="Similar to F6D8.5."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55622.1"
/db_xref="GI:5903063"
/translation="MASSRNSGSRKVGETIYISIPPTGVHKATVWLHDVGTGHC
VPLASLHPNKKWITVTPMRPVTSIGEVTTAMCDMPKSENNLDFENLYTNSF
ITSLSSEPDVIMKGVGIGLGAQALYTSVCFMGPVISPQIVIRINGMLPADDDVI
PSAFGKCADSLRMAGFPLFKQCGSKQLLYFVL1"
22422..24321
/gene="F6D8.8"
join(22422..22431,22558..22669,23113..23149,23253..23322,
23452..23556,23698..23830,24195..24321)
/gene="F6D8.8"
/note="Similar to F6D8.5."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55623.1"
/db_xref="GI:5903064"
/translation="MASGSINVSGLERQINTVYPTGIHKATVWLHDVGTGENSE
LPONLRPNKKWICTAPRRVTSIGETINMACDIAKVSENNDDGNTNVEKIT
SIFNSRPNVTKGAVGLGAAQALYTSVCFMGPVISPQIVIRINGMLPKMRKADS
LMAAFPTLTKQCGSKRLHLSRPFEDYFLNTH"
complement(25221..25968)
/gene="F6D8.9"
complement(join(25221..25673,25912..25968))
/gene="F6D8.9"
/note="Identical to gb|X91919 LEA76 homologue type1 from
Arabidopsis thaliana. ESTs gb|N97082, gb|Z27056 and
gb|Z29902 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55596.1"
/db_xref="GI:5903037"
/translation="NASHQEQSYRAGETRGKQKGTGEMGTGDKTOAKDQETA
QSAQKAEHTPAQAKDTSQAQTTQERAEKSDKTSVMEGTGEALKKKADALEYT
KETAEAGKERTSGILQGTGEQKQAMGATDAVHTLGLTFDEKKEHVSAPSTTTT
TTRETORK"
26849..27391
/gene="F6D8.10"
join(26849..26905,27022..27168,27251..27391)
/gene="F6D8.10"
/codon_start=1
/evidence=not_experimental
/product="F6D8.10"
/protein_id="AAD55597.1"
/db_xref="GI:5903038"
/translation="MSSSOQLSHSAGRYGVOLKKEEYLNNTSHAMNONADHTHSO
SOLSHEDDNPBLSIQASVITQGTGGVKNMAQADAVAKTLGMSFATNSPSSPAG
TTRNKGSKTI"
28814..30691
/gene="F6D8.11"
join(28814..28895,29040..29116,29376..29435,29835..30071,
30159..30221,30327..30395,30472..30517,30621..30691)
/gene="F6D8.11"
/note="Is a member of the Pf|00364 Biotin-requiring
enzymes family. ESTs gb|F19971 and gb|F19970 come from
this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55598.1"
/db_xref="GI:5903039"

```


/translation="MKYLSRTTLASVKAIOISTVPAETALADYKDSDEKSTVYN
THLMPSSEVALISEITDSSSIABEELKGGFRIVARLUTDESSPPQOIQVVA

Query Match 5.2%; Score 87.8; DB 8; Length 117737;

Best Local Similarity 53.3%; Pred. No. 3.2e-08;

Matches 185; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1110 TTGGCAGTGGGCCCTTACTTGGCAGCATTTGATGCTTGTTCAGAAATTCAGGGCCGT 1169
DB 115127 TTGACCATTTGATTTTGTTCGTTATTATTTTACCTTGATTTTTCACGGAGTT 115068
QY 1170 GGAGATTCCTTGGAGATTAACCATTTAGACATGGCATTTCACTGCCAGATTTGGCC 1229
DB 115067 GGATGAACATTTGTGTAGATGACATTCAGACATGCTTTATGACCTTGAGAGTGGGA 115008
QY 1230 ACGCTGCTGAGAGGACCAACGAGATTCCTCGACGAGCTCGACTGCTTTCATCTT 1289
DB 115007 ACGATGTCAGAGGACCAAGAGATGTTCAACACACATGACTGCTTACTTCACTCA 114948
QY 1290 GCTGGGAGAGAGGGCGCTTCTTTCGAAAGAAATGTCAGATTCGACAGCCAGCCGCC 1349
DB 114947 ATGGGGAGATCCGTTAAACAAAGATTAGCTCATTAATCTTGTATCTTACTTGAAGATCCA 114888
QY 1350 GAGCAGAGGCTCTCCATTCGAGGAGAAAGGGCTATGCGAGGATGTAAGAAATGTTTCATGAG 1409
DB 114887 GACGAGAGGCTACGATTTGGGGAGTGAAGACGCAAAATGTTGAGAAATGTAATGATGAG 114828
QY 1410 CACCATATGCTGAGAGATTCGCGCGCTGTGGAAGATGTCCTGAG 1456
DB 114827 CAACATATGTAACAGCGATTCGATGCTTCTGCTCAATGCATGAG 114781

RESULT 8

AC023010/c

LOCUS 147292 bp DNA linear HTG 18-APR-2000
DEFINITION Homo sapiens clone RP11-17A14, LOW-PASS SEQUENCE SAMPLING.

AC023010

AC023010.2 GI:7582616

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

1 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F., Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fensholt, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, U., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tittel, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.

Direct Submissions
Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 18, 2000 this sequence version replaced gi:6921557.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: U6614
Center clone name: 17_A_14

NOTE: This record contains 163 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 743: contig of 743 bp in length
* 744 843: gap of 100 bp
* 844 1646: contig of 803 bp in length
* 1647 1746: gap of 100 bp
* 1747 2505: contig of 759 bp in length
* 2506 2605: gap of 100 bp
* 2606 3370: contig of 765 bp in length
* 3371 3470: gap of 100 bp
* 3471 4280: contig of 810 bp in length
* 4281 4380: gap of 100 bp
* 4381 5149: contig of 769 bp in length
* 5150 5249: gap of 100 bp
* 5250 5964: contig of 713 bp in length
* 5965 6064: gap of 100 bp
* 6065 6832: contig of 768 bp in length
* 6833 6932: gap of 100 bp
* 6933 7656: contig of 724 bp in length
* 7657 7756: gap of 100 bp
* 7757 8512: contig of 756 bp in length
* 8513 8612: gap of 100 bp
* 8613 9417: contig of 805 bp in length
* 9418 9517: gap of 100 bp
* 9518 10329: contig of 812 bp in length
* 10330 10429: gap of 100 bp
* 10430 11212: contig of 783 bp in length
* 11213 11312: gap of 100 bp
* 11313 12081: contig of 769 bp in length
* 12082 12181: gap of 100 bp
* 12182 12894: contig of 713 bp in length
* 12895 12994: gap of 100 bp
* 12995 13738: contig of 744 bp in length
* 13739 13838: gap of 100 bp
* 13839 14539: contig of 701 bp in length
* 14540 14639: gap of 100 bp
* 14640 15395: contig of 756 bp in length
* 15396 15495: gap of 100 bp
* 15496 16245: contig of 750 bp in length
* 16246 16345: gap of 100 bp
* 16346 17086: contig of 741 bp in length
* 17087 17186: gap of 100 bp
* 17187 17926: contig of 740 bp in length
* 17927 18026: gap of 100 bp
* 18027 18818: contig of 792 bp in length
* 18819 18918: gap of 100 bp
* 18919 19695: contig of 777 bp in length
* 19696 19795: gap of 100 bp
* 19796 20570: contig of 775 bp in length
* 20571 20670: gap of 100 bp
* 20671 21447: contig of 777 bp in length
* 21448 21547: gap of 100 bp
* 21548 22302: contig of 755 bp in length
* 22303 22402: gap of 100 bp
* 22403 23163: contig of 761 bp in length
* 23164 23263: gap of 100 bp

23264 24028: contig of 765 bp in length
24029 24128: gap of 100 bp
24129 24872: contig of 744 bp in length
24873 24972: gap of 100 bp
24973 25721: contig of 749 bp in length
25722 25831: gap of 100 bp
25832 26599: contig of 778 bp in length
26600 26699: gap of 100 bp
26700 27460: contig of 761 bp in length
27461 27560: gap of 100 bp
27561 28341: contig of 781 bp in length
28342 28441: gap of 100 bp
28442 29199: contig of 758 bp in length
29200 29299: gap of 100 bp
29300 30083: contig of 784 bp in length
30084 30183: gap of 100 bp
30184 30928: contig of 745 bp in length
30929 31028: gap of 100 bp
31029 31786: contig of 758 bp in length
31787 31886: gap of 100 bp
31887 32640: contig of 754 bp in length
32641 32740: gap of 100 bp
32741 33464: contig of 744 bp in length
33465 33584: gap of 100 bp
33585 34353: contig of 769 bp in length
34354 34453: gap of 100 bp
34454 35198: contig of 745 bp in length
35199 35298: gap of 100 bp
35299 36061: contig of 763 bp in length
36062 36161: gap of 100 bp
36162 36924: contig of 763 bp in length
36925 37024: gap of 100 bp
37025 37800: contig of 776 bp in length
37801 37900: gap of 100 bp
37901 38663: contig of 763 bp in length
38664 38763: gap of 100 bp
38764 39531: contig of 768 bp in length
39532 39631: gap of 100 bp
39632 40387: contig of 756 bp in length
40388 40487: gap of 100 bp
40488 41273: contig of 786 bp in length
41274 41373: gap of 100 bp
41374 42145: contig of 772 bp in length
42146 42245: gap of 100 bp
42246 43022: contig of 777 bp in length
43023 43122: gap of 100 bp
43123 43864: contig of 742 bp in length
43865 43964: gap of 100 bp
43965 44694: contig of 730 bp in length
44695 44794: gap of 100 bp
44795 45464: contig of 670 bp in length
45465 45564: gap of 100 bp
45565 46349: contig of 785 bp in length
46350 46449: gap of 100 bp
46450 47230: contig of 781 bp in length
47231 47330: gap of 100 bp
47331 48102: contig of 772 bp in length
48103 48202: gap of 100 bp
48203 48981: contig of 779 bp in length
48982 49081: gap of 100 bp
49082 49867: contig of 786 bp in length
49868 49967: gap of 100 bp
49968 50741: contig of 774 bp in length
50742 50841: gap of 100 bp
50842 51608: contig of 767 bp in length
51609 51708: gap of 100 bp
51709 52475: contig of 767 bp in length
52476 52575: gap of 100 bp
52576 53113: contig of 738 bp in length
53114 53413: gap of 100 bp
53414 54164: contig of 751 bp in length
54165 54264: gap of 100 bp
54265 55003: contig of 739 bp in length

Query Match 3.9%; Score 65.8; DB 2; Length 147292;
Best Local Similarity 50.8%; Pred. No. 0.0018;
Matches 157; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 6 GCTGAGTTGATCCGACCCACAGTCTCTCTCGAGCCACCGCGTCCGACCGCGATG 65
DB 14464 GCCCGGGGGGAGAGGCGCCCGACGCCCGCGCGGAGCGGAGCCACCGCGG 14405
QY 66 GCAAGAGCCCGGTGCTTCGCGGTGCGGGGCGGAGGCGCGGCGGCTTCCACAC 125
DB 14404 GAGGGGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAC 14345
QY 126 CGGACCCAGCT 185
DB 14344 GGGGCGGGGCGG 14285
QY 186 TTCT 245
DB 14284 GGGCT 14225
QY 246 AAGACCG 305
DB 14224 CGTCT 14165
QY 306 GTCTCCCAT 314
DB 14164 GCCTGTAAT 14156

RESULT 9
AC109537/c 104486 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-230E2, *** SEQUENCING IN PROGRESS
DEFINITION *** 60 unordered pieces.
ACCESSION AC109537
VERSION AC109537.3 GI:21737934
KEYWORDS HTG: HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 104486)
AUTHORS Muzny D.M., Adams C., Adio-Oduola B., Ali-Goman F.R., Allen C.,
Alshrocks S.L., Amaralunge H.C., Are J.R., Ayale M., Banks T.,
Barbata J., Benton J., Bimaga K., Blankenburg K., Bonin D.,
Bouck J., Bowie S., Brivela M., Brown B., Brown M., Bryant N.P.,
Bunay C., Butch P., Burkett C., Burrell K.L., Byrd N.C.,
Carroll T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Davila M.L., Davis C., Davy-Carroll L., Decker D.A.,
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.T.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferriguto D., Flagg N., Ford P., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,

Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Hayak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B., Homel, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, J., Korvah, J., Kovar, C., Kratoch, J., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabba, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokem, S., Ogun, M., Okunola, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Qules, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Unpublished
Direct Submission
2 (bases 1 to 104486)
Worley, K. C.
Submitted (05-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 104486)
Worley, K. C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18846856.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GP2J
Center clone name: CH230-230E2
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 51833 bases at least Q40
Consensus quality: 55340 bases at least Q30
Consensus quality: 58186 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_diff_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 60 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1020: contig of 1020 bp in length
* 1021 1120: gap of unknown length
* 1121 2126: contig of 1006 bp in length
* 2127 2226: gap of unknown length
* 2227 3387: contig of 1161 bp in length
* 3388 3487: gap of unknown length

3488 5002: contig of 1515 bp in length
* 5003 5102: gap of unknown length
* 5103 6894: contig of 1792 bp in length
* 6895 6994: gap of unknown length
* 6995 8390: contig of 1396 bp in length
* 8391 8491: gap of unknown length
* 8491 9759: contig of 1269 bp in length
* 9760 9860: gap of unknown length
* 9860 10937: contig of 1078 bp in length
* 10938 11037: gap of unknown length
* 11038 12928: contig of 1891 bp in length
* 12929 13028: gap of unknown length
* 13029 14172: contig of 1144 bp in length
* 14173 14272: gap of unknown length
* 14273 15297: contig of 1025 bp in length
* 15298 15397: gap of unknown length
* 15398 16905: contig of 1508 bp in length
* 16906 17005: gap of unknown length
* 17006 18205: contig of 1200 bp in length
* 18206 18305: gap of unknown length
* 18306 19622: contig of 1317 bp in length
* 19623 19722: gap of unknown length
* 19723 20976: contig of 1254 bp in length
* 20977 21076: gap of unknown length
* 21077 22580: contig of 1504 bp in length
* 22581 22680: gap of unknown length
* 22681 24233: contig of 1553 bp in length
* 24234 24333: gap of unknown length
* 24334 25471: contig of 1138 bp in length
* 25472 25571: gap of unknown length
* 25572 26603: contig of 1032 bp in length
* 26604 26703: gap of unknown length
* 26704 28205: contig of 1502 bp in length
* 28206 28305: gap of unknown length
* 28306 30496: contig of 2191 bp in length
* 30497 30596: gap of unknown length
* 30597 32028: contig of 1432 bp in length
* 32029 32128: gap of unknown length
* 32129 34351: contig of 2222 bp in length
* 34351 34450: gap of unknown length
* 34451 36149: contig of 1699 bp in length
* 36150 36249: gap of unknown length
* 36250 37543: contig of 1274 bp in length
* 37544 37624: gap of unknown length
* 37624 38879: contig of 1256 bp in length
* 38880 38979: gap of unknown length
* 38980 40254: contig of 1275 bp in length
* 40255 40354: gap of unknown length
* 40355 41575: contig of 1221 bp in length
* 41576 41675: gap of unknown length
* 41676 42909: contig of 1234 bp in length
* 42910 43009: gap of unknown length
* 43010 44310: contig of 1301 bp in length
* 44311 44410: gap of unknown length
* 44411 45543: contig of 1133 bp in length
* 45544 45643: gap of unknown length
* 45644 46934: contig of 1291 bp in length
* 46935 47034: gap of unknown length
* 47035 48142: contig of 1108 bp in length
* 48143 48242: gap of unknown length
* 48243 49718: contig of 1476 bp in length
* 49719 49818: gap of unknown length
* 49819 51226: contig of 1408 bp in length
* 51227 51326: gap of unknown length
* 51327 52490: contig of 1164 bp in length
* 52491 52590: gap of unknown length
* 52591 53823: contig of 1233 bp in length
* 53824 53923: gap of unknown length
* 53924 55605: contig of 1682 bp in length
* 55606 55705: gap of unknown length
* 55706 57465: contig of 1760 bp in length
* 57466 57565: gap of unknown length
* 57566 58844: contig of 1279 bp in length

* 58845 58944: gap of unknown length
* 58945 61093: contig of 2149 bp in length
* 61094 61193: gap of unknown length
* 61194 62706: contig of 1513 bp in length
* 62707 62806: gap of unknown length
* 62807 64024: contig of 1218 bp in length
* 64025 64124: gap of unknown length
* 64125 65734: contig of 1610 bp in length
* 65735 65834: gap of unknown length
* 65835 67181: contig of 1346 bp in length
* 67181 69752: contig of 2372 bp in length
* 69753 71874: gap of unknown length
* 71874 71973: gap of unknown length
* 71974 73923: contig of 1950 bp in length
* 73924 74023: gap of unknown length
* 74024 76036: contig of 2013 bp in length
* 76037 76137: gap of unknown length
* 76137 77977: contig of 1841 bp in length
* 77978 80200: contig of 2123 bp in length
* 80201 80301: gap of unknown length
* 80301 82529: contig of 2229 bp in length
* 82530 84884: contig of 2255 bp in length
* 84885 86957: contig of 1973 bp in length

Query Match 3.8% Score 64.4; DB 2; Length 104486;
Best Local Similarity 53.6%; Pred. No. 0.0036;
Matches 155; Conservative 0; Mismatches 132; Indels 2; Gaps 1;

QY 23 GCCCAGTCTCTCTCTCGGAGCCGCGCTCGACCGGCGATGCGAAGACCCCGTCTT 82
Db 12630 GCGGACGTCGCTCCACGCGGCGCGCCCTCCGAGCGGCGCAATCGCGCGGCG 12571
QY 83 CGCGGTGGCGGCGGTGCGCGGCGCGCGCGGCTCACAACCGGACCCAGCTCTTCT 142
Db 12570 GCGGAGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12513
QY 143 CTTCTCTCTCTGCGCGCGGTGCGAGCTCGCATCGACAGAGGTTCTCTTCGCGGTC 202
Db 12512 CTTGCTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12453
QY 203 CTTGCGAGACCTTTCGACGCGCGGTGAGGAGCCCGCGCGCGCTCAACCGCGCGTGCAG 262
Db 12452 GCGGCGGCG 12393
QY 263 CGGAGTCCCTCGGCGTTTATGAGTTCAGAGTCTGCTGCTCTCTCTTC 311
Db 12392 CGGCGGCG 12344

RESULT 10
NCBI2N19 42273 bp DNA linear PLN 18-JAN-2002
LOCUS Neurospora crassa DNA linkage group II BAC clone B12N19.
DEFINITION A165987
ACCESSION A165987
VERSION A165987.1 GI:18376315
KEYWORDS Neurospora crassa.
SOURCE Neurospora crassa
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 Schulte, U., Aign, V., Hohseil, J., Brandt, P., Fartmann, B.,
AUTHORS Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 42273)
AUTHORS German Neurospora genome project.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik,

FEATURES
source
COMMENT
GSP-Forschungszentrum fuer Umwelt und Gesundheit, GmbH, Ingolstaedter Landstrasse 1, D-85764 Neuburg, FRG, E-mail: G.Mannhaupt@gef.de Project Coordinator: Ulrich Schulte, Institute of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf, E-mail: ulrich-schulte@uni-duesseldorf.de
BAC clone 12N19 (strain OR74A) is available at the Fungal Genetic Stock Center, <http://www.fgsc.net>
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany, <http://www.mwgda.com>
Information on performance of analysis and a more detailed annotation of this entry and other sequences of linkage groups II and V can be viewed at: <http://mips-gef.de/proj/neurospora>.
Location/Qualifiers
1. 42273
/organism="Neurospora crassa"
/db_xref="taxon:5141"
/chromosome="6"
1. 7212
/note="overlap to BAC B1B6, EMBL:AL442043, please refer to this entry for analysis and annotation"
9830..11707
/gene="B12N19.020"
join(9830..10364,10432..11357,11450..11707)
/gene="B12N19.020"
/note="differs in three residues when compared to PIR:T47219
contains amino acid permease signature (CGPASLIDFAIGCMFNVVYALGELAVMEP)
contains EST gb:BF39557, BF39423"
/codon_start=1
/product="amino acid permease NMAP1"
/protein_id="CAD21063.1"
/db_xref="GI:18376316"
/translation="MSDIKSGASGSHDKETRYDAATLPAYAGSVAAEQEDPMTRNG LNLSPFRPHFGDTEIERMTKSRHLNMLIGSIGAGFNGSKALAHNGPSLL IDPAICGMFNVVYALGELAVMEPVTGPTYSRFLIDPSKMGANVYVQAMITLPLELTVGCTGCMNDDTTWITTFVATITITINVEGTIGIEBFNMAAFKLGATIVG LLAELVAVGGKRSRDIHTGALMTSPGAFNANGFPGSCFVPAARFSGTELVG PFVAVNGRYKSLPGAIKQVFWRTTLFYLIGLFPVGLIIPYDENMLGANPIFNVAS RSNPLPSVLLIAFGPLAVNLAAGPTVFMILSLISGLAALFTWGSICLSIEFRK AMALKGHVEIIEPFAAGVWNGLGLVVLVLAQFYVAICPPGGGNTVEGFVQ YIALPVVIVFWIGFIMKRTSMLRVDQIDVDIGREHDMESINAYQALAAHPAKRN FHLFV"
9830..10364
/gene="B12N19.020"
/number=1
10365..10431
/gene="B12N19.020"
/number=1
10432..11357
/gene="B12N19.020"
/number=2
11358..11449
/gene="B12N19.020"
/number=2
11367..11389
/note="23 bp poly(c) tandem repeat"
11450..11707
/gene="B12N19.020"
/number=3
11995..14030
/note="36 bp atatatatatatatatatat tandem repeat"
14175..14909
/gene="B12N19.030"
14175..14909
/gene="B12N19.030"
/note="contains EST gb:BG279737, BG279736"
/codon_start=1
/product="putative protein"
/protein_id="CAD21064.1"
/db_xref="GI:18376317"
/translation="WGIIDYLRPRONGI PRKHTMPPDILSYQNDALDKRSHLSM

gene
24117..25698
/gene="B23H20..020"
join(24117..24204,24852..24890,25048..25062,25097..25177
25263..25297,25447..25518,25585..25698)
/gene="B23H20..020"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD21074.1"
/db_xref="GI:1876328"
/translation="MKVSMRGIIPSSQASIQRCGRDARKPESEAEAEEDCAS
KQASIMYELMLEIBSSSEARAGAGSEFFEDQDRKGALXQOTPAALASBGLA
ASIQDNEMLAEAVPDPASPTFGATTHKRVNADSGVSSVDSI"
24117..24204
/gene="B23H20..020"
/number=1
24205..24851
/gene="B23H20..020"
/number=1
24852..24890
/gene="B23H20..020"
/number=2
24891..25047
/gene="B23H20..020"
/number=2
25048..25062
/gene="B23H20..020"
/number=3
25063..25096
/gene="B23H20..020"
/number=3
25097..25177
/gene="B23H20..020"
/number=4
25178..25262
/gene="B23H20..020"
/number=4
25263..25297
/gene="B23H20..020"
/number=5
25298..25446
/gene="B23H20..020"
/number=5
25447..25518
/gene="B23H20..020"
/number=6
25519..25584
/gene="B23H20..020"
/number=6
25585..25698
/gene="B23H20..020"
/number=7
27953..30094
/gene="B23H20..020"
complement(join(27953..29206,29267..30094))
/gene="B23H20..020"
complement(join(27953..29206,29267..30094))
/note="similarly to ferroxidase precursor, YMR058W,
Saccharomyces cerevisiae, PIR:A55428
Contains Multicopper oxidase signatures
[GWLFGHIEHMEMISGLAVTF];Multicopper_Oxidase2
[HCHIEHMEMISGL]"
/codon_start=1
/product="related to cell surface ferroxidase"
/protein_id="CAD21075.1"
/db_xref="GI:1876328"
/translation="MKILISRIALLISITPAAHAAITVYDFNVSWITANDPGAEAPVI
GINWMPERRIDVQDRILVNLHNSLAGEDETSLHPFGLFNKGNSHNDGPVWVTCPI
PPASPTNFTVQDPTIVYHSHNKQVPDFSRGLVHDPKAPFEYDEELVITVSDG
YHRSNLSIOPRFKSNPTGAEPVPSALMNDQNTMPMDVDRDTYLFRIANGIFAG
QYIVWEDITTVEMDGVYHPTETNMVYVAAQRCSFLITAKKETSRRNPPIIAMDQ
TLEDTIPGLNTAVSYLLVNPAPLPLEPTDELILPDLDTLIPPDNLTLTSPDQT
ITLVNTWYMLASANYAEFNDIITYVDPRPVPTLVTVISPNSSSSGNDNLTTPDSIVET

YTHPFLSHNSIVQVILNLDGSRHPIHLHGHQFQIILHRSAGDEGTWSPSSSSSTS
 SFTSNGSHHTAPMRDITLVEPENGNAVIRFADNPQVWLFCHIEWHMISGLATFI
 EAPSVQOQTINLPADHLAACRRODIPPEGNAAGRGKDKNTLSQWILVGGSPSPK
 PIRAGFTGGIINALTMSCTFGIIGVAVVWVGFSPGSENTAAAVTKSPAAAGSS
 GSGSVYTVTVDSKAGSASVVRKGTAVSESLRFSGEERAVNGVGMATSKKDG
 VVVDVAITRA*

exon
 complement (27953..29206)
 /gene="B23H20.030"
 /number=1
 complement (29207..29266)
 /gene="B23H20.030"
 /number=1
 complement (29267..30094)
 /gene="B23H20.030"
 /number=2
 /number=2
 /note="24 bp agtcagcagtc tandem repeat"
 30220..30243
 31765..33093
 /gene="B23H20.040"
 /note="31765..33093"
 /gene="B23H20.040"
 /note="similarity to high-affinity iron permease catr2,
 candida albicans, TRMBB:AF195776.1
 contains EST gb:AA901925, BF072546"
 /codon_start=1
 /product="related to high-affinity iron permease"
 /protein_id="CAD21076.1"
 /db_xref="GI:18376330"
 /translation="MQLRSVPILVFRRTIYSVILVFKOTLASPSSSSSPA
 EATSHAAATRAADDELKHPSSSTNNSSPOOPVVDIAYVNAIPROIIIGTV
 LGFLTAIGAVICVFTTLGVNKNSSSAGELNMGATCLIASVIGVALLRVER
 MKEKWRKMIKAMEGKRNKSKGKGVAGGEDMERETVEKRWKVKAMMERI
 VMFLPFTVLREVEALVAVGSPAPATSIPALVIGISLVGVVTKGSSST
 KLQFLVLSLILVAAGLFSRAIWALESQMAVIGSVAELSGSPSYIDKSW
 HVDGSPVNGGGWIGIFNAIFGWTNATVGSVIGYDYLWLVIVSFLVMRYEVKSK
 WPLIPKTEVEDGSKGSLFGRLSMRKRA"

exon
 31765..32576
 /gene="B23H20.040"
 /number=1
 32577..32636
 /gene="B23H20.040"
 /number=1
 32637..33093
 /gene="B23H20.040"
 /number=2
 33444..35567

intron
 32577..32636
 /gene="B23H20.040"
 /number=1
 32637..33093
 /gene="B23H20.040"
 /number=2
 33444..35567

gene

Query Match 3.4%; Score 58.2; DB 8; Length 122151;
 Best Local Similarity 50.9%; Pred. No. 0.079;
 Matches 138; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

1168 GTGAGAAATGCTTGAAGATTAATTAAGCAATGCAATTCAAGTTGCCAGTATTGG 1227
 13728 GTAAAGAACTTTTGTATCTGCGCTTGAAGCAGTTAAAGGGCGCTGCCGTGTGG 13787

1228 GCAAGGCTCTGAGGGGACCAAGGAGATGCTCTGAGCGCTCGACTGCGCTTGTGATC 1287
 13788 CGGCAATTAATGATGAGCCGACGAGGAGCGTGTGAGGAGCGGGAGTGTAGAG 13847

1288 CTGTGGAAGAAGAGGGGCGTGGCCCTTTGCAAGAACAATCTCGACATCGCAAGCAGC 1347
 13848 ACCGAAATGAGTAGGGGAGTGGGCTAAGATTAGATTAAGGTTTGAATGGATGGGGG 13907

1348 CCGAGACAGAGGCTCTCATATGGGGGAAAGGGCTATGCGCAGGTGAAGAAATGTTATCG 1407
 13908 AGGAAAGCTTGAAGAGATGGAAAGAAAGGGGTGAGAGGGGTCAAGGGTAGTTGGCG 13967

1408 AGCACCAGATGCTGAGAGGATCGCGCGGT 1438
 13968 ATACGAGATGCGGAGAGGTTTGAAGAGAT 13998

AC126306/c
 LOCUS
 DEFINITION
 AC126306
 VERSION
 AC126306.1 GI:21699124
 KEYWORDS
 HTG; HTGS_PHASE1
 SOURCE
 Rattus norvegicus
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 152607)
 Munz D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
 Albrooks S.L., Amaralungu H.C., Are J.R., Ayale M., Banks T.,
 Bardola J., Benton J., Bimase K., Blahndung K., Bonini D.,
 Bouck J., Bowie S., Briteva M., Brown E., Brown M., Bryant N.P.,
 Bunay C., Burck P., Burkett C., Burrell K.L., Byrd N.C.,
 Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
 Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
 Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
 Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
 Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,
 Doucwaile K.U., Draper H., Dugan-Rocha S., Durbin K.U.,
 Eathart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
 Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
 Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
 Garrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
 Hernandez O., Hodgson A., Hognes M., Holloway C., Hollins L.E.,
 Homsi F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.B.,
 Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
 Karlsson E., Kelly S., Khan U., King L., Koryah J., Kovar C.,
 Kratochvic J., Kuresh A., Landry N., Leal B., Lewis L.C., Lewis L.,
 Li J., Li Z., Lichtege O., Lieu C., Liu J., Liu W., Louissege H.,
 Lozato R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
 Maheshwari M., Mapua P., Martin R., Martindale A., Mattheiz E.,
 Massey E., Maehney E., McLeod M.P., Meador M., Mei G., Metzger M.,
 Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
 Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
 Nguyen N., Nickerson E., Nwokenwo S., Ogun M., Okunnu G.,
 Oreguine N., Oviedo R., Pace A., Payton B., Peery U., Perez L.,
 Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
 Rivers M., Rojas A., Rojebokan I., Rolfe M., Ruiz S., Savary G.,
 Scherer S., Scott G., Shen H., Shooshitari N., Stinson I.,
 Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H.,
 Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,
 Tansey J., Taylor C., Taylor T., Telirod B., Thomas N., Thomas S.,
 Usmani K., Vasequez L., Vera V., Villalon D., Vinson R., Wang Q.,
 Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
 Williams G., Williamson A., Wleczek R., Wooten S., Worley K.,
 Wu C., Wu Y., Zhou J., Zorrilla S., Nelson D.,
 Weinstock G., and Gibbs R.

REFERENCE
 2 (bases 1 to 152607)
 Worley K.C.
 Direct Submission
 Submitted (05-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 152607)
 Worley K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information

Center project name: GZ1Y
Center clone name: CH230-207N13
----- Summary Statistics -----
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; Version 0.990329
Consensus quality: 92958 bases at least Q40
Consensus quality: 99800 bases at least Q30
Consensus quality: 103479 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1163: contig of 1162 bp in length
* 1263: gap of unknown length
* 1283: contig of 1143 bp in length
* 2406: gap of unknown length
* 2506: contig of 1264 bp in length
* 3770: gap of unknown length
* 3869: gap of unknown length
* 5212: contig of 1343 bp in length
* 5213: gap of unknown length
* 5312: contig of 1628 bp in length
* 5313: gap of unknown length
* 6941: gap of unknown length
* 7041: contig of 1065 bp in length
* 8106: gap of unknown length
* 8206: gap of unknown length
* 8206: contig of 1153 bp in length
* 9359: gap of unknown length
* 9459: contig of 1817 bp in length
* 11275: contig of 1388 bp in length
* 11276: gap of unknown length
* 11376: contig of 1388 bp in length
* 12764: gap of unknown length
* 12863: gap of unknown length
* 14193: contig of 1330 bp in length
* 14194: gap of unknown length
* 14294: gap of unknown length
* 15902: contig of 1608 bp in length
* 15902: gap of unknown length
* 16002: contig of 1652 bp in length
* 17654: gap of unknown length
* 17753: gap of unknown length
* 18822: contig of 1069 bp in length
* 18823: gap of unknown length
* 18923: gap of unknown length
* 19959: contig of 1037 bp in length
* 20060: gap of unknown length
* 21674: contig of 1615 bp in length
* 21675: gap of unknown length
* 21774: gap of unknown length
* 22857: contig of 1183 bp in length
* 22958: gap of unknown length
* 23057: gap of unknown length
* 24259: contig of 1202 bp in length
* 24359: gap of unknown length
* 24360: gap of unknown length
* 25974: contig of 1615 bp in length
* 25975: gap of unknown length
* 26075: gap of unknown length
* 27682: contig of 1508 bp in length
* 27683: gap of unknown length
* 29108: contig of 1426 bp in length
* 29209: gap of unknown length
* 31375: contig of 2167 bp in length
* 31376: gap of unknown length
* 32514: contig of 1039 bp in length
* 32515: gap of unknown length
* 32515: gap of unknown length
* 33698: contig of 1084 bp in length
* 33699: gap of unknown length
* 33799: contig of 1516 bp in length
* 35314: gap of unknown length
* 35415: contig of 1574 bp in length
* 36988: gap of unknown length
* 37089: contig of 1276 bp in length
* 38365: gap of unknown length
* 38465: contig of 1795 bp in length
* 40259: contig of 1795 bp in length

* 40260: gap of unknown length
* 40360: contig of 1650 bp in length
* 42010: gap of unknown length
* 42010: contig of 1473 bp in length
* 43583: gap of unknown length
* 43583: contig of 1205 bp in length
* 44888: gap of unknown length
* 44888: contig of 1083 bp in length
* 46071: gap of unknown length
* 46171: contig of 1631 bp in length
* 47802: gap of unknown length
* 47802: contig of 1944 bp in length
* 49846: gap of unknown length
* 49846: contig of 1306 bp in length
* 51252: gap of unknown length
* 51352: contig of 2031 bp in length
* 53382: gap of unknown length
* 53383: contig of 2084 bp in length
* 55667: gap of unknown length
* 55667: contig of 2128 bp in length
* 57794: gap of unknown length
* 57794: contig of 1425 bp in length
* 57895: gap of unknown length
* 59420: gap of unknown length
* 59420: contig of 2266 bp in length
* 61866: gap of unknown length
* 61866: contig of 2667 bp in length
* 64453: gap of unknown length
* 64453: contig of 1869 bp in length
* 66422: gap of unknown length
* 66422: contig of 2248 bp in length
* 68770: gap of unknown length
* 68770: contig of 2549 bp in length
* 71419: gap of unknown length
* 71419: contig of 1875 bp in length
* 73394: gap of unknown length
* 73394: contig of 1796 bp in length
* 75290: gap of unknown length
* 75290: contig of 3042 bp in length
* 78432: gap of unknown length
* 78432: contig of 2931 bp in length
* 81463: gap of unknown length
* 81463: contig of 2322 bp in length
* 83885: gap of unknown length
* 83885: contig of 3060 bp in length
* 87045: gap of unknown length
* 87145: gap of unknown length
* 89459: gap of unknown length
* 89459: contig of 2315 bp in length
* 91921: gap of unknown length
* 91921: contig of 2362 bp in length
* 92022: gap of unknown length
* 94573: gap of unknown length
* 94574: gap of unknown length
* 94674: gap of unknown length
* 98616: gap of unknown length
* 98616: contig of 3064 bp in length
* 101679: gap of unknown length
* 101679: contig of 3064 bp in length
* 101680: gap of unknown length
* 101779: gap of unknown length

Query Match 3.3%; Score 57; DB 2; Length 152607;
Best Local Similarity 48.6%; Pred. No. 0.15;
Matches 156; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
* 19 CCGAGCCGACAGCTCTCTCGGCGCCACCGCTCCGACCGGACGCAAGACCCGCT 78
* 40025 CCGGCGCGGCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39966
* 79 CCGTTCGCGGCG 138
* 39965 CCGTTCG 39906
* 139 TCGCTTCG 198
* 39905 CCG 39946
* 199 GCG 258

Db 39845 TCTCTCCCCCCCCCTGCTCCCGGCGCTCTCGCGGCTCTCGCGCCCGCGCCCGCCCCCGCGCG 39786
Qy 259 CCAAGGAGAGTCCCTCGGCTTCAAGAGTCAAGTCTGCTCTGCTCTCCCAAGAGC 318
Db 39785 CCGCGCGCGCGCGCCCCCCCCCTCCCGCGCGCTCTCTGCGCGCCCCCGGAGGC 39726
Qy 319 TCTCCCTCTGCTGCTGCTGAC 339
Db 39725 CCGCGCGCTCTCCCGCGCCCC 39705

RESULT 13
AB026653/c 35762 bp DNA linear PLN 27-DEC-2000
LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MWC8.
DEFINITION AB026653 BA000014
ACCESSION AB026653.1 GI:4757409
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MWC8.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
P1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480
2 (bases 1 to 35762)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:yakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MWC8
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli1.zool.iastate.edu/cgi-bin/bp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MSJ11 and the 3' clone is MSJ1.
Location/Qualifiers
1. 35762
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MWC8"
/clone_lib="Mitsui P1"
join(2370..2410,2893..3040,3127..3389,3462..4095)
/note="gene_id:MWC8.1"
/codon_start=1
/evidence=not experimental
/product="protein kinase-like protein"
/protein_id="BAB02873.1"

/db_xref="GI:9294592"
/translation="MQLLNSCCGKGFDEKKKEKESMVFSLKELHAATNSFNYNK
LGEGRFGSVYWGQLMDSQIAVRLKEWNSNEEIDFAVELEILAIRKULHLEVRVC
AEGORRLVLEYEMONLSIVSHLHGOHSAECLMDTMRKMTIASQALAIYHLHAPH
IYHGVRAVSNTLLDSEFAEAVTDFGKGLMDDDDGGAATKASNNGYISPCDAGSK
ESPTSDVYSRGLILMTVYSKRPRLRLNPTRTITEMVLPVYRNNGEIVKRLSE
BHARKLKVVLVGLMACOTDPRKPTMSRVEMLVNSEKESISLEANNPLFNNPVSS
NENRBNHVAEESDVILIEDHQQOQE"
complement(join(4337..4534,4726..4953))
/note="gene_id:MWC8.2
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02874.1"
/db_xref="GI:9294593"
/translation="MSSGVVPEGMALCYPTLSRSSVHNSLEPTKMTMSVKNPVSF
RNSNSCGSKRSFRVMASDRSDQNDGQIGDELTNLKLTIGVGAQAALIKGSVL
PQITRPNLTLALPILIIAPVISVILLSSSSSKKQN"
join(9979..10477,10563..11050)
/note="gene_id:MWC8.3
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02875.1"
/db_xref="GI:9294594"
/translation="WKYSSRRELRTQRYIVNPVMTWMSIPTREBSILPIGLSN
GSGKYYVVKLTATTPSQLKVYVDSKGGQKRFYCIHPWSSSMPVTCQAVENK
ALHMLHMDPIVAPNPKYRACIIHRSQEMHSHGGAIVSGLPIQYRMQYHLLIOL
VCYFPNDHHLISWTLVDYEGSLMKHSDVFRDWSGIVPIODMYRMQLLETLQ
MSFLMVFMDKRAPETPRKRIIPMLTCHPNMLLVLYLPSIVSLDVATKTLQI
TKIKKCTCPKRGESGYMVPMLTLEBSLSDHERLLECHRTYVR"
complement(join(11749..12006,12087..12305,12418..12660,
12743..12907,13038..13744,13916..14024,14098..14184,
14277..14375,14479..14589,14667..14936))
/note="emb|CAA16573.1
gene_id:MWC8.4
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02876.1"
/db_xref="GI:9294595"
/translation="NNLPHADLSLDPRYNSGPPGEPILSHRTLSPGPFQGEEDYR
RNTSYHDANREBDTDSRASPPIHDSRSLPFGMDMSPPRLHGGDTWPHDR
TGSYCVTPSVWVDLPKSVSDPAVYVQVALISDPGITSALVPLGNDVPLEYS
IKKEFPKSLPQAPPKKILMMRNQTLLERRCSLEDMNLLSDIDISMLITPLE
LEAAVSYNDEYQTEDETGDIPLPTTISDVGSSTYIDHDSADETSNASTM
KHENALKTLPSRNSTADVNTVMHLELYGLDSSPQEKYERLSTNGDPAATGY
TRGISGGLORLDGSDRKRQRLTESIKKHVSDPEASTAYEPLVNOGMDIHGE
AKGMVGAAGCTEPOXDIAIVOSREPRKIKYVITLTKRLTAKADTELISRLNO
ELARQPLSTRDLAEVLETTESCKQMKETUIDEKERTIOXMDRELKQCMEM
ESFLNSIKOEKHTITANESLVOENQMLQIINDIFENFERKHELELYAKAEIK
VLVEVYSIRTTQSDRLQELSGIMKEKEMERIVOREKDEEPAKADKULHRCVU
ONRLOCNVFPDIEEGKLIMDSLSSEALIELATVSNRGLLIATQILSEVKIK
LTSGRGRCDDLVRRKMLEVLIIDNARLRQVNSVLRCSLGHGISVREKGEVDEBGS
SIDLARTVMSKILEK"
join(15425..17406,17640..17745)
/note="gb|AAC35225.1
gene_id:MWC8.5
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02877.1"
/db_xref="GI:9294596"
/translation="MILNRPKSPFNSLSIFKALLMSTITESINDYSRFLISLVC
KTTDOPKLOHSQITRGVAPNPFOKKLFVFWCSLGGHVSAYVKLFVKIPEDVTV
NNMIGKMSKYDQDGEQVLYMMIKGVYPDSTPEPLINGLRDGLACGKILHCH
VKEGSLSNLYQNLYQMSLGLMDMARGVPPRCCEPVEFSMNLGSMKMEKE
ESTIELBEMERLVSPVTILLVLASCKVXKXDLCKRHYEVSCRTSPSLRBN
LVNVAACGEMDIARIFRSKARDAVSWTSIVKQYERNGALATFTDQFPRDL
SWTMMIDGYLRAGCFNBSLEIFRMOASAMIDPEYRVSVLTCALVSLDEWIKT
YIDKRIKNDVVVGNALLIDMYFKGCSSEQAQGVFMDNDKRTFAMVAVGLANNGQ
OEAIKVFPQMDMSIQPDITLYGLVSAENHGWADQARFKFKMSDEIRLEISVHY
GQWMDMGRAGLVKEAYEIRIKRPMNPNISIVGALLASRLHDEPMALAKKILEL

CDS

EPDNGAVALLGNIVAGCKRWKDLREVRKRIYDVAIKKTPGSLIEVNGFAHFAVD
KSHLSEELIYMKLELAQESTFAAIVLPDTSLLFPAQKXLEHSHQVTLTKNPIESI
NMFEVALFMKA"
complement(join(18580..19367,19641..19665))
/note="emb|CAB10501.1
gene_id=MVC8.6
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02878.1"
/db_xref="GI:9294597"

/translation="MTSLKAVRCOTVRIOPRSFHRIDRYALGYDKNRHKILAFVD
HLAEGHVFGEIYDPSSSMRVIDYTPMDIQYQGVSLNNTYFPAQKLTDEE
VEEPDELLYDPSRTERPEPLPLPHFYLETYSLSVSEDEALVLRDLSYELD
INLTITERSAIVSNFVKDMKPLTGFOFMDGSGFIDEEKKVAVFDTLTKKQTV
TKRQTVYIIGQDQFKVSLGAPLADPHELCPLVCPSPASPLVQLVN"
complement(join(20658..21062,21115..21180))
/note="gene_id:MVC8.7
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02878.1"
/db_xref="GI:9294598"

/translation="WRIVSIQYSTRSLIHIMELNNSVVDNHSNISEWYDQNV
HMEYNDADQKVCISSVLKRSLVIAIIFIVSTVTSLSKLTALKNYIIHS
NRLIAGRSSCPILKTSMDGIFNICKIQVAKKIANSTVALVVMRRR"
complement(join(21613..21855,21947..22003,22111..22946,
23211..24168))
/note="gene_id:MVC8.8"
/codon_start=1
/evidence=not experimental
/product="glycosyl transferases-like protein"
/protein_id="BAB02880.1"
/db_xref="GI:9294599"

CDS

/translation="MEDLRSLPKLGSFRSSLSGRSTPRGSLTKRVHSGRTPRREG
GSGCAVQFRSNRLIYLLITLITLTYLGFYQSHADNDNVEELRFGKLEDDIHA
VEQNKRDVLADESSHAVDHNTIYHLGVNRMHVTIAKKEGSTRSRSVPRRTKA
SRSTRIRSTQVKRYMETKELDEQDELINIVYGLRPGSLDEPLIEMSK
RSGTCRSDPKRLVMSRRRFLVFLHESMTGAPISMELASLLSCGTYAYVLESR
CGILQETRRRIKVEDKESLSTKTMKADYVIGASVCSNITDQYMDHPRGSSDIA
MYMERRREYDRAKPYLDKVLILFSESVKMLWCEDHVAIRQPIYPLSVN
DELAFVAGVSSLNTPLTLOETMKERQKLESVTEFGITKDLVMSLSINPKG
QLLLLESVALIEREQEOQVARNOSKIINLKLIGKELISLNRHRSRRKKKI
TSFVADNHSVALTSATGRKLLSGNVTOKODLKLIGSVSGSNVAAYEMSKPSI
NGLNSVLTWPATTRVASYADVYVNSQGVETGRTIEMAYGAPVIGTQAG
GTKEIVENVTGLHPVGRAGNKKVLAONLFLLRNPSTRLOGSGGREIYERKYMCH
MYKRPDIYVKMRP"
complement(join(25764..26471,26569..26672,26774..26803,
26889..26948,27028..27060,27161..27211,27291..27335,
27450..27497,27584..27652,27751..27798,27884..27949,
28035..28082,28174..28239,28333..28380,28465..28539,
28636..28683,28773..28835,28927..28974,29061..29126,
29227..29274,29369..29434,29524..29571,29658..29732,
29856..29925,30039..30154,30249..30290,30368..30422,
30517..30591))
/note="gene_id:MVC8.9
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB02881.1"
/db_xref="GI:9294600"

CDS

/translation="MGTKFLALGLSLICLVSPFOVSCODEGTGSLSTLIDIEHXOT
SYNSLQNEAVDQETESQKNSVTDNNTISLSISEPALLETIKESYASBLGAVTD
EYDKSSMLDIELEFBAHINELKAGSDGINVRSKXODEBARRRHKMLEALRFE
AAHAGFEOLKTDDSAOGLDDQSAROSMDEIFRFEATYKGLGKQKADLTGLINDE
EHAAROKMLETERFERPEATYKGLGELRHSSTTDEQASAROKMDEITERFEAT
SGIKELKNATYKDVDDKQDAROSMDALERFEAVTESRQLEDIADKAGAD
DQSAROSMDEITERFEATYKGLINLDFSEGDSSAESARRNSMLEALIRERFEAA

CDS

Query Match 3.3%; Score 56.4; DB 8; Length 35762;
Best Local Similarity 53.7%; Pred. No. 0.17;
Matches 117; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Oy	1221	GTATTGGGACCGGCTGCTGGAGGACCAAGCATGCTCCTGACGGCTCGACTGGCCCTT	1280
Db	21855	GTGCTTGGAACAGATGCTGGAGAACAAAGAGATGAGACATGTTACAGGGCTTA	21796
Oy	1281	CTGCATCTGCTGGAGAGAGGCGTGGCGCCTCTTCAAGAAACATCGTAGACTGCA	1340
Db	21795	CTTCATCTGTGGGAGGAGGAGGTAAACAAAGTTTGGACAGAACTTTTGTCTTCTT	21736
Oy	1341	AGCAGCGGACAGAGGCTCTCCATGGGAGAAAAGGCTATGGCAGGGTAAAGAAATG	1400
Db	21735	AGAAACCATCTTCAAGAGGCTCAACATGAGTACCAAGAGCTAAATGCTTGAAGAGATG	21676
Oy	1401	TTGATGAGCACCACATGCTGCTGAGAGAGATCGCGCGGT	1438
Db	21675	TACATGACGACATGTCACAGATTTTGATGCT	21638

RESULT 14

LOCUS	AF128457	70311 bp	DNA	linear	PLN 25-APR-1999
DEFINITION	Oryza sativa subsp. indica BAC clone 16f19 phg20725 region,				
ACCESSION	AF128457				
VERSION	AF128457.1	GI:4680335			
KEYWORDS	complete sequence.				
SOURCE	Oryza sativa (indica cultivar-group).				
ORGANISM	Oryza sativa (indica cultivar-group).				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Euharoidae; Oryzaceae; Oryza.				
REFERENCE	1 (bases 1 to 70311) Llaca, V., Lou, A., Young, S. and Messing, J. Microcollinearity in cereal genomes Unpublished				
AUTHORS	2 (bases 1 to 70311) Llaca, V., Lou, A., Young, S. and Messing, J.				
REFERENCE	Submitted (15-FEB-1999) Waksman Institute, Rutgers University, 190 Submitted (15-FEB-1999) Waksman Institute, Rutgers University, 190 Fellinghuysen Rd., Piscataway, NJ 08854-8020, USA				
JOURNAL	location/Qualifiers				
FEATURES	1..70311				
source	1..70311				

1. 70311
/organism="Oryza sativa (indica cultivar-group)"
/cultivar="Teging"
/sub_species="indica"
/db_xref="taxon:39946"
/chromosome="10"
/map="25.4"
/clone="16f19"
/note="phg20725 region
putative map location based on mapping of Oryza sativa EST
clone E10310 6Z found in GenBank Accession Number C9155"
join(162..367,452..556)
/note="predicted gene - truncated (3'); similar to Oryza
sativa EST clone 97AS2194 found in GenBank Accession
Number AA153140"
join(<2846..3003,4043..4325,4560..5015,5301..5833,
6229..6424,6527..6685,6834..7670)
/product="hypothetical protein"
join(2846..3003,4043..4325,4560..5015,5301..5833,
6229..6424,6527..6685,6834..7670)
/note="similar to Oryza sativa EST clones S1328_1A found
in GenBank Accession Number D39744 and 96BS0679 found in
GenBank Accession Number AA753512"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD21627.1"
/db_xref="GI:4680336"

/translation="MDGEARWLAAGVAGADLVAAALROLEFLAAYDRRRMYEG
PLERAIHREVILKWCIFQVOYLKDKCRLYGIIIDNNVSSIRASRQSKWMAAO
YKPEPELNTSSDNGSYANAGAADISYDVAAYKQSSFFQYDTTHMDQRFLE
BALARKGLYIKTNQSRKMLRIVTIDVVMHSHQHPATYCHDLKIGRIVE
HDDTDQRSKGLKDTGSGTGFQFNAGARAKAMRGNLPSPVNSPQMSOE
VNEFSVGRASQITLITETVIELFLQIVDIKNLPPAIPRENYIWTNKPDMFISD

CDS

misc_feature

mrna

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

RNA
complement(join(25162..25820,26472..227849))
/product="putative receptor_kinase"
CDS
complement(join(25162..25820,26472..27849))
/note="similar to Oryza sativa EST clone E2866_2A found in Genbank Accession Number C73087"
/codon_start=1
/product="putative receptor_kinase"
/protein_id="RAD27636.1"
/db_xref="GI:4680345"
/translation="MYTLLLAFLRSTLFLLLLAAGAVDPGAAAPDTAAAPDTA
AALLRLASLIDPTNALEAMSPSPSPCDDEARWPRVOYCNVGLGLALANLSGD
EPBALRLGLGASINLIRNNSGSLPASLAIVRSILALYLSNAPSGPVDVPAAM
SMUKGLYLDRNNSGSELDPAGAIAPRLDDELHLDNRTEGVSKLPATLLPVSIN
RTVLGEVAANAENFESAPACNPGOLCAPSGSGACAAAPGRHSMAMPMSAADYPA
VOETLSAFVQVIGIMLVLLVAGAMVIMLRDEGTSTASGTEHPAIGAPSGNLSVPI
AAGAAASAPVQVITMEQSGSGGVGAGARQVAEVLVMSNAEPFLPELMAKSAEVL
NGLTIGSAGYAARAKNRVTVAVKRMIDNRVGAEEHEHIRMGLBETPTSPSSAYIT
ARKSSGSPSDOSPRVVLVDWPMRRLAVGVAGLSYHEKLGIPAMRLVSMGTADFD
APPPEPRLGKNGNLTLDALIEPRLVDGPEPLVNTSQAPHAMAFERSPEASAGCA
GGAARRALASASRDYVDCIGIVLLELYTGKPSQGYLTARGGDVQMAAFAVAGT
EEBVDVPAVGAAGPRAARLLRVGRTCTTPEPSRPSMADVARNVQVAGGAS"
25893..26223
35340..35604
/rpt_type=dispersed
35968..37755
/note="similar to Oryza sativa receptor kinase-like protein gene found in Genbank Accession Number U72724"
/product="hypothetical protein"
complement(join(39112..39523,42036..42349))
/codon_start=1
/product="hypothetical protein"
/protein_id="RAD27634.1"
/db_xref="GI:4680343"
/translation="WRKSPSTSPVYAAAVLLMYTLMAGVGEAAVLPPSCNPT
LITPAGTLFGGVGPACQALRAQACIYASPNVGSYISPNARLFTVGLG
IRSKTNCCKNRDSTSHLRVHDSNSTATSSAVAPRPPILAAGTGRLVHD
RRRLDRPLRSPSPRLAAASATASSTGSRVLVDHLRLTTPPLVNDCHRSS
TAKRLSACADELMPHAKHSL"
39637..39788
/rpt_family="putative MITE C"
39737..39788
/rpt_type=dispersed
40957..41072
/rpt_type=dispersed
41907..42271
/rpt_family="putative MITE B"
/rpt_type=dispersed
41907..42271
/note="putative LTR"
43355..43549
/rpt_family="putative MITE D"
/rpt_type=dispersed
join(45393..45921,46479..46533,47382..47453,47458..47671)
/codon_start=1
/product="hypothetical protein"
/protein_id="RAD27633.1"
/db_xref="GI:4680342"
/translation="MCLMRRRTAAEMVDGAGAEKYLSTRGVAAALBEGQMAPEIL
AHRHSRRPLDAPPRDRRAHIFAPRLDEPLIHRHRTDITMGSGPGRKSSS
GEHSGSPFRMLLEVVDGRRGHFTVRLRDCGRRRGGCAPTAOLMGLYARRGRG
RRGAAGSTGVGAGADAPLRKVFLLYQSDWRRAAYALSISGTLNSAPR
SVSAAGCTGMLAMPSPPHORQPPPPRRQPPSPPKRIVAEITYTNDDIVITI
IILIIIVFYVIRG"
46305..46456
/rpt_type=dispersed
47149..47522
/note="putative LTR"
complement(47668..48017)
/rpt_family="putative MITE C"

```

RESULT 15	LOCUS	AF119222	77605 bp	DNA linear	PLAN 26-APR-1999
	DEFINITION	AF119222			
	ORIGIN	Oryza sativa	BAC clone 1.H19,	complete sequence.	
	FEATURES	AF119222			

Eunariotia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzoae; Oryza.

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (pages 1 to 77605)	Llaca, V., Lou, A., Young, S. and Messing, J.	Microvnteny in a gene-dense region in maize, sorghum, and rice	Unpublished	2 (pages 1 to 77605)

RECEIVED
DATE
TITLE
JOURNAL
Submitted (10-JAN-1999) Wakeman Institute, Rutgers University, 190
Frelinghuysen Rd., Piscataway, NJ 08855-8020, USA
108855-8020, USA

[illegible][illegible]

```
mRNA
complement(<29032..29790)
/product="hypothetical protein"
complement(29032..29790)
/codon_start=1
/product="hypothetical protein"
/protein_id="A027673.1"
/db_xref="GI:4680493"
/translation="MAVDGEMPALESQFNOCAITAEBAALPDSTEPPTNVASODE
ESGAGESSESDDSDDEVIDIIEGDDGSDADYTGTPLIEGMLITTTVDKI
LALRRRFPMPATGGERLMMVSPFEFEELIAEIAAALVDSQKPECCAMIAEK
RHPEGAVVEETSEWMTTICRPFGRVSEDDDDLLCDVDANECKCGDDMAN
EFIDEEDDSCSDVDEEBEEDKSCCGRMVRLR"
complement(30165..30488)
complement(join(31774..32072,32911..33076))
/product="hypothetical protein"
complement(join(31774..32072,32911..33076))
/codon_start=1
/product="hypothetical protein"
/protein_id="A027683.1"
/db_xref="GI:4680503"
/translation="MLFOARQPTAYLNSEYSDKRTIMWYSMLLGAARSSSEF
RLQSRRLQVAYPRFTGSLNHSQCFYTVLGLVHMLPEVQLPKSADMAIGDVS
LSRQVIRLMLIDRPAIRGGGEPHSHRVHGVYVCVAHSHLHVLAL"
join(<34690..34730,35085..35261,35329..35549,
38631..39499)
/product="hypothetical protein"
join(34690..34730,35085..35261,35329..35549,38631..39499)
/codon_start=1
/product="hypothetical protein"
/protein_id="A027674.1"
/db_xref="GI:4680494"
/translation="WQSVADSLPHLIKOTISPPFTKIAODTINLVPARLPFKRTSL
RLROKRTIFGTGTLGPAPKIRHPQASFTLPSPPSLHSSGTEGGSGHGRRA
ATRGAVSTEAVRVEVATAGAGGVRLDEPVRBEPASGQPHSPPLTLAT
HATTRIRPHSSASTAGIRPLSLSTCIGDRAIRIIAPLHCTVFSRSARAAA
AAAGDLDPDVHARPARLRDQAPHPDQPGGRAGAAGDGDLLGAAGR
RRRLHVGAAARREVEVRLPRDELBEHDAVAVGRRGALGSGAGAGGGGG
CLGRPEREHGVRRLRVEDREAVAHVADGLOVGVODVAGLEVARRRRRREVEGAS
HAQPHRRDQLIVRQAHADGDPHRRPICHPLRALVT"
complement(join(<40016..40032,40110..40588,40596..41539))
/product="receptor kinase-like protein"
complement(join(40016..40032,40110..40588,40596..41539))
/note="similar to Oryza sativa EST clone E2886_2A found in
Genbank Accession Number C73087"
/codon_start=1
/product="receptor kinase-like protein"
/protein_id="A027675.1"
/db_xref="GI:4680495"
/translation="WTLILAFRLSTLFLILAGAIANDPGAAAEDTAAPDTENAA
ILRLKASLIDPTALBAMSPSPCDETHRMPRVQCYNGVLGLRLALNSGPD
FAALSRLEPGLHSINLIRNNFSGPLASIAVLRALYLSRNASGFPDGVFAMSW
LKULYLRNDFSGELPAGAIAGAPRLOELHDNRLEGVPSKLPATLRFLVNSHRL
TGVLEAVARFNESAFAAGNPLCGAPSGACAAAGPAPHTAMPNMGADYPAVQ
BETSFTVMGIIIMLVYLIVAGAMVLMRODEGTASGTPPLGAPLRALCAPPAAG
AAASAOIYTWBGGSGAGAGGAGVGVGAGAROVAFVIMSNAGBEGLPELMKASAE
VLAGNCTLSAYKAAARNGTAVYKRMAMRVGRALFEHITMLGELRHNVLSPIGY
HTRKEKLIASEMPERGSLLVHLGKISF"
complement(join(<42779..42931,43886..44595,45422..45637,
45851..45988,46130..46266,47466..47502,48912..49126,
52727..53003,53024..53135))
/product="putative sugar transporter protein"
complement(join(42779..42931,43886..44595,45422..45637,
45851..45988,46130..46266,47466..47502,48912..49126,
52727..53003,53024..53135))
/codon_start=1
/product="putative sugar transporter protein"
/protein_id="A027676.1"
/db_xref="GI:4680496"
/translation="WAPRIYDTPACAHGVORLAQTALFAHYKXRYOLEKHCIN
DDGEVVTINVAGVRCRAACVLDGGRRRGGGAGATVDCPDADAVRGAGAVW
GAGAAGMLCTAAGTCGPMVRTEPOLCSSTPSTNOSAGNOALSPSPSHRLTHFP
RALPIPALARSKLASPRVRGTGAPSFRRGAGAAARIIPSEVPSDRPNHGRRSNY
LRRALVLDMMKSTVTSVAVNSIGTYLLGMDFTTVLGNPSAANIHKKEFGILNNGS
```

```
Query Match          3.3%; Score 56; DB 8; Length 77605;
Best Local Similarity 53.1%; Fred. No. 0.22;
Matches 119; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
IDGILLAVSVEGSIATVSEGSLLDMJGRRIVYSYCLIFMMTLIPSPNMIMTALFA
PSLVEFALLVLYLSEPRMLVSDGKISBARISLQMLRGKDVSDRSAGEIALADGN
MITETAVGHAHVAGVRSQSFTGNTNOMSRHSTPYWHLSDPVLDLGSIHSESLGA
GRSYSPVFNSEFNIVBOETSBOGRNDSLQOSREAVSABEGNNNDNLQASLLSIVASA
ETNDINTSPTEBSSSYLRHRHGTSGLAQDLISLSLHDHIDEBDEEIHIMALSQPA
90 GCGCGGTGCGCGGAGCGCGCGGCGGCTTCAACAACGACCGAGCTCTCTCTCTC 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19574 GTGAGCTGCGCGCGCGCGCGCGCGCGCGCTGAGCTCCCGCGAGCTTACCTC 19633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150 CTCGCGCGCTCGCGAGCGCTCCGCAATCCAGAGAGGTTCTCTCGCGGAGTCC 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19634 ATGCGCGCGCGCGCTTCTTCAACATGTGCTGAGAGTCTTCTTCAACCACT 19693
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 GACCTTGCAGCGCGCGTGGGAGACCGCGCGCTCAACACCGCGCTGCGCAAG 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19694 ATCTTCCCGACCGCGCGCATCAACCCCGCCACCTCAACGACCTCGAGCA 19753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 CCGCTCGGTTCAATGAGTCCAGAGCTGCTGCTCTCTCCCA 313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19754 CTTCTCATGTTCTTCTCTTGGCATCTCGCCTCTCTCTCCA 19797
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 4, 2003, 21:37:10
Job time : 299 seconds

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 14:31:05 ; Search time 283 seconds
(without alignments)

13543.820 Million cell updates/sec

Title: US-09-938-294-2

Perfect score: 1702
Sequence: 1 atcgcgtcgcagttgtatcc.....gagsgatctcaatcttagc 1702

Scoring table: IDENTITY NUC
Gapop 10.0 , Capext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq, 101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	589.6	34.6	1715	21 AAC42068	Arabidopsis thaliana
2	154.6	9.1	284	21 AAA31140	Plant microsatelli
3	49.8	2.9	912	21 AA233717	HSV-1 VP22 cellular
4	48.4	2.8	1076	20 AA230007	Optimised nucleoti
5	48.4	2.8	1082	24 ABA2912	Orphanophshae hy
6	48.2	2.8	856	20 AA219784	Herpes simplex vir
7	48.2	2.8	856	20 AA207807	HSV-1 tegment pro
8	48.2	2.8	900	21 AA29385	HSV-1 VP22 polypep
9	48.2	2.8	912	20 AA26227	HIV-1 VP22 polypep

10	48.2	2.8	912	21 AA288468	
11	48.2	2.8	912	22 AAF58996	HSV-1 VP22 coding
12	48.2	2.8	912	24 ABA93386	HSV-1 VP22 protein
13	48.2	2.8	950	19 AAV17085	Herpes simplex vir
14	48.2	2.8	1225	22 AAK1456	VP22/57 fusion con
15	48.2	2.8	1257	24 ABK1810	DNA encoding pcDNA
16	48.2	2.8	2004	22 AAD09259	VP22-Cre fusion DN
17	48.2	2.8	2055	22 AAD09268	VP22-CreStrepTag fu
18	48.2	2.8	2247	22 AAD09280	VP22-F1pe fusion D
19	48.2	2.8	4488	22 AAD09286	pT7-VPCS vector DN
20	48.2	2.8	6404	21 AAC61514	Nucleotide sequenc
21	48.2	2.8	6420	21 AAC61515	Nucleotide sequenc
22	48	2.8	65140	22 AAD17186	Streptomyces nous
23	48	2.8	125401	22 AAT80413	Tyactone synthase
24	46.8	2.7	43280	18 AAV19862	Gallus domesticus
25	46.6	2.7	954	19 AAV19862	Gallus domesticus
26	46.6	2.7	1347	19 AAV19864	Gallus domesticus
27	46.6	2.7	1587	19 AAV19863	Gallus domesticus
28	46.6	2.7	2233	19 AAV19861	Gallus domesticus
29	46.6	2.7	5027	19 AAV19860	Gallus domesticus
30	46.4	2.7	21721	20 AAX83427	Human lipolysis st
31	46.4	2.7	23187	21 AAS50273	Human leptin fragm
32	46.4	2.7	23187	22 AAF62311	Human leptin fragm
33	46.2	2.7	4403765	22 AAI99683	Mycobacterium tube
34	46.2	2.7	4411529	22 AAI99682	Mycobacterium tube
35	46	2.7	1905	18 AAT84863	Aspergillus niger
36	45.6	2.7	2653	21 AAD00338	Wheat rathhouse sy
37	45.6	2.7	44861	24 AAS20000	DNA encoding pyrid
38	45.2	2.7	1551	22 AAD05562	Human secreted pro
39	45.2	2.7	1590	24 ABO54274	Human ovarian anti
40	44.6	2.6	27541	22 AAD17185	Streptomyces nous
41	44.6	2.6	4403765	22 AAI99683	Mycobacterium tube
42	44.6	2.6	4411529	22 AAI99682	Mycobacterium tube
43	44.4	2.6	1216	21 AAC65527	Zea mays DNA fragm
44	44.4	2.6	1440	22 AAI59765	Human polynucleoti
45	44.4	2.6	1440	22 AAI59766	Human polynucleoti

ALIGNMENTS

RESULT 1	
AAC42068	
ID	AAC42068 standard; DNA; 1715 BP.
XX	
AC	AAC42068;
XX	
DT	17-OCR-2000 (first entry)
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 34167.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
XX	
KW	Protein identification; signal transduction pathway;
XX	
KW	metabolic pathway; promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126284.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.

PR	19-APR-1999;	99US-013007.
PR	21-APR-1999;	99US-010449.
PR	23-APR-1999;	99US-010450.
PR	23-APR-1999;	99US-010891.
PR	28-APR-1999;	99US-011449.
PR	30-APR-1999;	99US-0112048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0112486.
PR	06-MAY-1999;	99US-0112487.
PR	07-MAY-1999;	99US-0112863.
PR	11-MAY-1999;	99US-0134236.
PR	14-MAY-1999;	99US-0114218.
PR	14-MAY-1999;	99US-0114219.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135184.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136382.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138090.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140981.
PR	30-JUN-1999;	99US-0141827.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142820.
PR	12-JUL-1999;	99US-0142877.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144336.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144811.
PR	21-JUL-1999;	99US-0145086.
PR	22-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145222.
PR	26-JUL-1999;	99US-0145291.
PR	27-JUL-1999;	99US-0145913.
PR	28-JUL-1999;	99US-0145913.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146388.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	05-AUG-1999;	99US-0147300.
PR	05-AUG-1999;	99US-0147192.
PR	06-AUG-1999;	99US-0147300.
PR	06-AUG-1999;	99US-0147300.
PR	09-AUG-1999;	99US-0147418.
PR	09-AUG-1999;	99US-0147493.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17	


```

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 34.6%; Score 589.6; DB 21; Length 1715;

Best Local Similarity 68.1%; Pred. No. 2,2e-156;

Matches 818; Conservative 2; Mismatches 382; Indels 0; Gaps 0;

```

OY 266 GACTCCCTCGGGTTCATGAGTCCAAAGCTGCTCTCTGTCCTCCATGAGCTTCCT 325
DB 351 GAATCCTCTAGATTTCATGAAATCGAAGCTGTTTGTAGATCCCAAGCTATCACT 410
OY 326 CTCTGGTGGTCCACTTTTACTGATGATGATGATGATGATGATGATGATGATGATGAT 385
DB 411 CTGATGTSAGCTTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 470
OY 386 AGTGTGTGATTAACAACAGAGATCACAGAAAACAAGATGATCAATATAGCTTGA 445
DB 471 CGTTGTTTGATCACTATCAAAAGCACTGGAAGATGAAAGTTGCTATAGCTTGA 530
OY 446 GCATAGGATGTAACATGAGTGCAGGTTTACAGCTAGAGAGAGAGAGAGAGAGAG 505
DB 531 ACACAGAGTGTGATGAGAGATGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAG 590
OY 506 TATTGCTTAAAGCTGATGCTGTTATCTTAAACACAGCTGCTGCTGCTGCTGCTG 565
DB 591 TACATCTCTCAAGGCTGATTTGATTTGCTTAAATAGCTGCTGCTGCTGCTGCTGCT 650
OY 566 CCCGTCTGAAAGATCATGTTCTTAAAGCTTCCGAAAGATTTTGTGGTGCATCGA 625
DB 651 TCGTGTCTTAAAGAAATGTTGTTAAAGTTCTTAAAGTCTTAAAGTCTTAAAGTCT 710
OY 626 AATGCTGGCATTTACTTAAAGTGAATATGCAACATCTTCTTTGTTGCTGAGC 685
DB 711 GATGAGAGACACTTTTCAATGCTATTTGCTCAAAACCTCCCTTTGTTGAGGAGC 770
OY 686 CATGATTGATCTCATACAGGCTAGTATTTGAAATGAGAGAGAGAGAGAGAGAGAG 745
DB 771 CATGATTGATCGATGACAGCTGATATCGAAAAACAAGACCAAGCTCGCTAGG 830
OY 746 AATACAGATCCCAAACTTATGTTGTTCACTGGGGAATGTTAAAGATCTAAAGAG 805
DB 831 GATTAAATGCCCAAACTTACGTTGGCAGCTCAGAAATGCAAGGAAATTAATGAGAG 890
OY 806 TGCTGAGACAAATGTCGAGAGAGAGCTTACAGGGAACATATTCGTAATCCCTTGA 865
DB 891 AGCTGAGATAGTGTGCAAGAGAGATTCCTGCTAGACCTCGAAGATCTCTTGA 950
OY 866 ACGAGTGAAGATCTCTGTTTGAATTAATAACAGTATCAAGAGAGAGAGAGAGAG 925
DB 951 GCGGATGAAGACTTACTCTTTGGCATTAATATGATGATCGGTGGAAGAGCCAGAG 1010

```

```

OY 926 CTATTTCTTCAAGCATTTTATCAGGCTTTCAGCTCATTCACACAGAACTTAAGT 985
DB 1011 TCTGTTCTCCAGAGCTTCCAGAAAGTCTTGAAGAAATCAAGAGAGAAATCTCAGGT 1070
OY 986 GCTTGAATACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1045
DB 1071 MCAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1130
OY 1046 TCAGTTACGTAATTTGTTGTAAGAAACAGATTCATACCTGCTCAATTTTGTAAACA 1105
DB 1131 AGAGCTACGCAACTTGTTCGAGAAAGAACTTGAAGAACTTGTCCACTGCTAACAA 1190
OY 1106 GACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1165
DB 1191 AACTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1250
OY 1166 CCGTGAAGAAATGCTTTGTAAGATTAACATTAAGCAATGCAATTCAGTTGCAATTT 1225
DB 1251 CAGAGAGAGAAATGCTTTGTAAGATTAACATTAAGCAATGCAATTCAGTTGCAATTT 1310
OY 1226 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285
DB 1311 TGTATCTGACAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1370
OY 1286 TCTTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1345
DB 1371 TAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1430
OY 1346 CCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1405
DB 1431 AGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1490
OY 1406 GAGAGACCAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1465
DB 1491 GGAATCATATGTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1550
OY 1466 GG 1467
DB 1551 GG 1552

RESULT 2
AAA31140
ID AAA31140 standard; DNA, 284 BP.
XX
AC AAA31140;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #101.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
XX
XX WO9967421-A1.
XX
XX PD 29-DEC-1999.
XX
XX PF 25-JUN-1999; 99WO-NZ00092.
XX
XX PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukala JV, Bloksberg LN, Giem W.
XX
XX WPI, 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for

```

the detection of polymorphic genetic markers -

Claim 1; Page 106; 392pp; English.

Sequences AAA31040-432093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

Sequence 284 BP; 97 A; 48 C; 73 G; 66 T; 0 other;

Query Match 9.1%; Score 154.6; DB 21; Length 284;

Best Local Similarity 74.0%; Pred. No. 1.8e-33;

Matches 196; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

568 CTGTTCTGAAGATCATGTTCTTAAGCTCTCCGAAGTTTGTGTGATCATGAAA 627
 20 CTAGCTCAAGACAAATGTTCTTACTGTTCTCCAAAATTTGTGTGATTCACGAAA 79
 628 TCCGTGGCACTTACTTAAGTTGAATATGCAACATCTTCCCTTTGTGTGAGCA 687
 80 TTAGAGGCGATATTTCACAACTGAGATATGTCAGACCTTATGTGCGACGGGCA 139
 688 TGATTGATTCATACAGCGCTGATTTGGAAATGCGACATGCGATCGCTGAAA 747
 140 TATTGATTTCTCAACACAGCTGAATCTGGAAGCCAGAACTCAAGAACTGTGGGGA 199
 748 TACAGATGCCCAAACTTATGTTTTCACCTGGGGAATGAAAGAACTAATGAAATTG 807
 200 TCAAGATGCTTAAGATTTATGTCGTCACCTTGAATATGCCAATGATTAATGAAATG 259
 808 CTGAAGACATGTCGCAAGAGAT 832
 260 CAGAAATATGTGCGCAAAAGGT 284

RESULT 3

AA293717

ID AA293717 standard; DNA; 912 BP.

XX AA293717;

DT 16-NOV-2000 (first entry)

DE HSV-1 V22 cellular localisation signal coding sequence.

XX Ubiquitin ligase; SCF; F-box protein; targeted degradation;
 KM destabilisation; proteolysis; drug discovery; gene therapy; cancer;
 KM oncoprotein; Huntington's disease; gene knockout; delivery systems;
 ss.

OS Synthetic.

XX Herpes simplex virus-1.

FN WO200022110-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-US23705.

PR 09-OCT-1998; 98US-0103787.

XX (HARD) HARVARD COLLEGE.

XX Zhou P, Howley P;

XX MPI; 2000-317970/27.
 DR P-PSDB; AA83261.

Targeting degradation of polypeptide useful for treating cancer and
 other proliferative disorders, involves conjugating polypeptide with
 ubiquitin protein ligase or inhibiting ubiquitination using organic
 compound

Disclosure; Page 75-76; 185pp; English.

The F-box proteins are a family of ubiquitin ligases (SCF ubiquitin
 ligases) which can be used for the targeted degradation of a target
 polypeptide in vivo. Targeted degradation is achieved by expressing
 the ubiquitin ligase in a cell linked to the interaction domain of
 the target polypeptide and thereby recruiting the target polypeptide
 to the ubiquitin ligase. Such methods are useful for decreasing or
 increasing the level of a target polypeptide and for creating and
 expressing a destabilized polypeptide which is subjected to SCF
 mediated proteolysis. Degrading any desired protein in a cell is
 useful for preventing or treating diseases caused by the presence of
 abnormal amount of the specific polypeptides, for drug discovery and
 for gene therapy. Diseases treated include cancer, by degradation of
 oncoproteins, Huntington's disease, other proliferative disorders and
 microbial infections. The method provides a quick and easy
 alternative to gene knockout technology. The target polypeptide can
 be degraded at all stages, or a specific stage, of development in the
 mature animal. The hybrid ubiquitin ligase may also include an
 optional localisation sequence such as the HSV-1 V22 sequence.

Sequence 912 BP; 149 A; 353 C; 289 G; 121 T; 0 other;

Query Match 2.9%; Score 49.8; DB 21; Length 912;

Best Local Similarity 47.6%; Pred. No. 0.0018;

Matches 147; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

3 TCGGCTCGAGTTGATCCGAGCCCAAGTCTTCTTGGGCTCACCGGCTCGACGGCG 62
 325 TCGGAGGGGCGGACCGACACCAACACCGGCTCCCGGGGCCCCGAAACCAACGGGTTG 384
 63 ATGGCAAGACCCCGATGTTGCGGTGGGCGGTCGCCGAGCGCGGGCGGTTTAC 122
 385 GCGACTAAGGCCCCCGGCGCGCGGCGGCGGAGACACCGCGGCAAGAAATCGGCCCA 444
 123 AACCGGACCCAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 182
 445 CCAAGATCCCGGCACTTCCAGACGCCCCCGGTGACGGCGCAACCGGATCCAGACA 504
 183 GGGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 242
 505 CCGCGCAGGGGCTGGCCAGAAAGCTGACATTAGACCGCCCCCAAAACCCGGAAGGG 564
 243 CTGAACAGCGCGGTGCGACAGCGGAGTCCCTCGGGTTTCAATGAGTCCAACTGTGCTC 302
 565 CAGTACACCCCGGGGTGGCGGCTTTAAAGCGGCTTCTTGTGCGCGGTGGGCGCC 624
 303 CTGCTCTC 311
 625 CTGGCGGCC 633

RESULT 4

AA230007

ID AA230007 standard; DNA; 1076 BP.

XX AA230007;

DT 26-JAN-2000 (first entry)

DE Optimised nucleotide sequence encoding organophosphorus hydrolase.

XX Barley; alpha amylase leader sequence; plant expression;
 KM organophosphorus hydrolase; OPH; environmental detoxification;

KM detoxification: organophosphorus neurotoxin contaminated soil; pesticide;
 KM insecticide; fungicide; herbicide; chemical warfare; transgenic plant;
 KM ss.

OS Synthetic.
 OS Pseudomonas diminuta.
 OS Hordeum vulgare.

PH Key Location/Qualifiers
 FT CDS 3..1076
 FT /tag= a
 FT sig_peptide 3..71
 FT /tag= b
 FT /note= "barley alpha amylase sequence"
 FT mat_peptide 72..1073
 FT /tag= c
 FT /note= "Pseudomonas diminuta organophosphorus
 FT hydrolase sequence"

PN W09953037-AZ.
 PD 21-OCT-1999.
 XX 14-APR-1999; 99WO-US08202.
 PF 15-APR-1998; 98US-0081846.
 XX (PROD-) PRODIGENE INC.
 PA
 XX Jilka JM;
 PI
 XX WPI: 1999-620418/53.
 DR P-PSDB: AAY43487.

XX Novel nucleotide sequence, used to detoxify organophosphorus compounds
 PT used as e.g. insecticides, fungicides, herbicides or chemical warfare
 PT agents

XX Claim 6; Fig 4A-B; 44pp; English.

XX The present sequence represents an optimised nucleotide sequence
 CC encoding an organophosphorus hydrolase (OPH) protein attached to an
 CC optimised sequence barley alpha amylase sequence. The products and
 CC constructs can be used for environmental and in situ detoxification of
 CC comprising exposing contaminated material to OPH, or detoxification of
 CC organophosphorus neurotoxin contaminated soil. The products can also be
 CC used as agricultural and domestic pesticides, including insecticides,
 CC fungicides and herbicides or chemical warfare agents. The OPH can also
 CC be used to hydrolyze other chemical agents or compound containing P-O,
 CC P-S, P-CN and P-S bonds. The products can be used to produce transgenic
 CC plants which are capable of expressing OPH at levels greater than that
 CC of the typical 10-15mg/l of E. coli or baculovirus.

XX Sequence 1076 BP; 149 A; 484 C; 276 G; 167 T; 0 other;

Query Match 2.8%; Score 48.4; DB 20; Length 1076;
 Best Local Similarity 53.1%; Pred. No. 0.0049;
 Matches 103; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 124 ACCGAGCCAGCTCCTCCTCCTCGTGCCTGAGCCGATCCAGATCCAGACAG 183
 DB 16 ACTGAGCCTCTCCCTCTCTCTGCTCTCGGCTCTCGGCTCTCTCGGACGGCA 75
 QY 184 GATTCTCTCTCGGCGGTCCTCTGAGACCTTTCGAGCGGCGGAGACCCCGCGCC 243
 DB 76 CCGGCGACCCGATCAACACCGTGCAGGCGGATCAACATCTCGAGGCGGCTTCAACC 135
 QY 244 TCAACACCGCGTCCGAGCGGAGATCCCTCGGAGTTCATGAGTCCAGCTCTGCTCC 303
 DB 136 TCAACACGAGACATCTGCGGCTCTCGGCGGCTCTCGGCGCTGCGGAGTTCT 195
 QY 304 TCGTCTCCATGAG 317
 ||| ||||| ||

DB 196 TCGGCTCCGCAAG 209

RESULT 5

ABA92912
 ID ABA92912 standard; DNA; 1082 BP.

XX ABA92912;

XX 09-APR-2002 (first entry)

XX Organophosphate hydrolase gene (OPH, E.C. 3.1.8.1).

KM Organophosphate hydrolase; enzyme; dimeric metalloenzyme; fungal; nickel;
 KM metal cofactor; gene; plant; E.C. 3.1.8.1; zinc; cobalt; manganese; ds.

XX Flavobacterium BP.

XX W0200196543-AZ.

XX 20-DEC-2001.

XX 14-JUN-2001; 2001WO-US19174.

XX 15-JUN-2000; 2000US-211732P.

XX (PROD-) PRODIGENE INC.
 XX (GEMV) GENENCOR INC.

XX Hood E, Howard JA, Bailey M, Van Gastel FJC, Ward M, Wang H;
 XX Woodard S;

XX WPI: 2002-090204/12.

XX Improving recovery of active enzyme e.g. laccase, which requires
 PT transitional metal cofactor e.g. copper for activity, from a plant, by
 PT introducing plant nucleotide sequences encoding the enzyme and exposing
 PT it to cofactor

XX Example 6; Fig 20; 81pp; English.

XX The present invention describes a method for improving the recovery of
 CC an active enzyme from a plant where the enzyme requires a transitional
 CC metal cofactor for activity. The method comprises introducing into the
 CC plant nucleotide sequences encoding the enzyme and exposing the enzyme
 CC to the metal cofactor. The method is useful for improving recovery of
 CC active enzyme which requires a transitional metal cofactor for activity,
 CC preferably for improving recovery of active laccase which requires
 CC copper for activity. The method can be used for improving recovery of
 CC active organophosphate hydrolase (OPH, E.C. 3.1.8.1) which requires
 CC zinc, nickel, cobalt or manganese for activity, where the method further
 CC comprises adding bicarbonate ion salt. The present sequence encodes OPH
 CC isolated from Flavobacterium sp., which is used in an example from the
 CC present invention.

XX Sequence 1082 BP; 151 A; 484 C; 278 G; 169 T; 0 other;

Query Match 2.8%; Score 48.4; DB 24; Length 1082;
 Best Local Similarity 53.1%; Pred. No. 0.0049;
 Matches 103; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 124 ACCGAGCCAGCTCCTCCTCCTCGTGCCTGAGCCGATCCAGATCCAGACAG 183
 DB 16 ACTGAGCCTCTCCCTCTCTCTGCTCTCGGCTCTCGGCTCTCTCGGACGGCA 75
 QY 184 GATTCTCTCTCGGCGGTCCTCTGAGACCTTTCGAGCGGCGGAGACCCCGCGCC 243
 DB 76 CCGGCGACCCGATCAACACCGTGCAGGCGGATCAACATCTCGAGGCGGCTTCAACC 135
 QY 244 TCAACACCGCGTCCGAGCGGAGATCCCTCGGAGTTCATGAGTCCAGCTCTGCTCC 303
 DB 136 TCAACACGAGACATCTGCGGCTCTCGGCGGCTCTCGGCGCTGCGGAGTTCT 195

XX Stratford LU, Patterson AV, Kingsman SM, Kan O, Griffiths L;
 PI Mitrophanous K;
 XX WPI; 1999-540852/45.
 DR P-PSDB; AAY27404.
 XX
 PT New producting activating agent targeted to selected cells or tissues,
 PT particularly hypoxic cells, for treating e.g. tumors or inflammation
 XX
 PS Example 7; Fig 3D; 149pp; English.
 XX
 CC The invention provides a new producting activating agent that comprises: (i)
 CC a localization domain (LD); other than a tumor-selective antibody and a
 CC producting activating domain (PAD); (ii) at least one nucleic acid encoding
 CC a cytochrome P450 and under control of at least one constitutive or
 CC inducible expression control sequence or (iii) a modified hematopoietic
 CC stem cell (MHSC) containing at least one nucleic acid encoding a PAD and
 CC under control of elements as in (ii). The producting activating agent or
 CC vectors that express them, are specifically used to treat tumors,
 CC inflammation, atherosclerosis and muscular dystrophy, but may also be
 CC used to treat many other conditions, e.g. cerebral malaria, rheumatoid
 CC arthritis, or conditions associated with hypoxia, hypoglycemia, or
 CC ischemia, or to deliver antibiotics, antiviral agents, analgesics,
 CC anesthetics, anti-inflammatories, antineoplastic agents and diagnostic
 CC agents. LD optimize activity of PAD, e.g. by delivering it to selected
 CC locations or by delivering it to neighboring cells (bystander effect),
 CC and allow a reduction in dose of producting and thus of systemic side-
 CC effects. Nucleic acids encoding the agent may be expressed selectively
 CC in hypoxic cells. The present sequence represents a DNA encoding the
 CC HSV-1 tegument protein VP22 (complement sequence 106391 to 105486 of
 CC Genbank sequence X14112). This is used in the construction of a fusion
 CC protein comprising VP22 and a human P450 reductase derivative alP450R.
 XX
 SO Sequence 856 BP; 137 A; 339 C; 270 G; 110 T; 0 other;

Query Match 2.8%; Score 48.2; DB 20; Length 856;
 Best Local Similarity 47.2%; Pred. No. 0.0049;
 Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTCGAGTTGATCCGAGCCCAAGTCTCTCTCGGAGCCAGCCGCTCGAGCCGAGC 62
 DB 322 TCCGAGAGGGGCGGAGCGACACCCAGCCGCCCCGGGGCCCCGGAACCCAGCGGGTG 381
 QY 63 ATGGCAAGACCCCTGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122
 DB 382 GCGACTAAGGCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 441
 QY 123 AACCGAGCCAGCT 182
 DB 442 CCGAATTCGCGCGCACTCCAGACGCCCCCGGCGTGAAGCGGCGCAACCCGATCCAGACA 501
 QY 183 GGGTCT 242
 DB 502 CCGCGGAGGGGCTGCGCAAGAAAGCTGATTAAGTACCGCCCCCAAAATCCGACGCG 561
 QY 243 CTCAACACCGCGCTGCGCAAGCGGAGTCCCTCGGAGTTCATGAGTCCAAAGCTCTGCTC 302
 DB 562 CCATGACGCCCGCGGAGTGGCGGCTTAACAAGCGCGTCTGCGCGCGGTCGGGCGCG 621
 QY 303 CTGCTCTCC 311
 DB 622 CTGGCGGCC 630

RESULT 8
 AAA29395
 ID AAA29395 standard; DNA; 900 BP.
 XX
 AC AAA29395;
 XX
 DT 12-SEP-2000 (first entry)
 XX

DE HSV-1 VP22 polypeptide coding sequence.
 XX
 XX hbs72; telomerase; catalytic subunit; reverse transcriptase; life-span;
 KW retinoblastoma; p53; tumour suppressor; inhibitor; arteriosclerosis;
 KM proliferation; immortal; tumour therapy; macular degeneration; activator;
 KM INK4; HSV-1; VP22; fusion protein; ss.
 XX
 OS Herpes simplex virus 1.
 XX
 FH Key Location/Qualifiers
 FT CDS 4..894
 FT /*tag= a
 FT /product= VP22
 XX
 XX MO200031238-A2.
 XX
 XX 02-JUN-2000.
 XX
 XX 24-NOV-1999; 99WO-US27907.
 XX
 XX 25-NOV-1998; 98US-0109891.
 XX 17-FEB-1999; 99US-0120549.
 XX
 XX (GENE-) GENETICA INC.
 XX
 XX Hannon GU, Beach DH;
 XX WPI; 2000-400055/34.
 XX P-PSDB; AAY96574.
 XX
 PT New method for increasing the proliferative capacity of cell lines
 PT comprises administering agents reversibly activating telomerase
 PT activity and reversibly inactivating Rb/INK4 and/or p53 pathways useful
 PT in treating age related diseases
 XX
 XX Disclosure; Page 31; 123pp; English.
 XX
 CC The HSV-1 VP22 polypeptide can be fused to a retinoblastoma (Rb)
 CC inactivator protein sequence to aid targeting and internalization.
 CC The invention concerns methods and reagents for extending the life-span,
 CC e.g. the number of mitotic divisions, of a cell. The method relies on
 CC activation of a telomerase activity and inhibition of one or both of a
 CC Rb/INK4 pathway or a p53 pathway. Phosphorylation of Rb by
 CC cyclin-dependent kinases, cdk4 and cdk6, releases the cells into the
 CC division cycle. Binding of INK4 family members, e.g. the tumour
 CC suppressor p16INK4a, inhibits kinase activity and results in growth
 CC arrest. Rb inactivators can selectively and reversibly inactivate an
 CC Rb/INK4 pathway, especially an Rb/p16INK4a pathway. The oncoprotein MDM2
 CC is a cellular inhibitor of Rb/E2F function and the p53 tumour suppressor
 CC and can also be used in the methods. Other molecules which can be used
 CC include cdk4 or cdk6 mutants. In particular, a cdk4 mutant is one which
 CC differs from at one or more of residues K22, R24, H95 and/or D97.
 CC Additional constructs include a papilloma virus E7 protein, or other
 CC viral oncoprotein which bypasses Rb and/or p53. Antisense constructs of
 CC the Rb and p16INK4 genes may also be used. The methods are useful for
 CC increasing the proliferative capacity of cells. The cells are
 CC subsequently of use in pharmaceutical and cosmetic preparations used to
 CC treat conditions related to (premature) ageing, e.g. macular degeneration
 CC and arteriosclerosis. The cells can also be used to replace tumour cells
 CC lines in vitro and for studies on biochemical and physiological aspects
 CC of growth and differentiation. Long lived (immortal) cells could also be
 CC of use in the production of normal or genetically engineered
 CC biotechnology products.
 XX

SO Sequence 900 BP; 144 A; 354 C; 282 G; 120 T; 0 other;

Query Match 2.8%; Score 48.2; DB 21; Length 900;
 Best Local Similarity 47.2%; Pred. No. 0.0051;
 Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTCGAGTTGATCCGAGCCCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 62
 DB 313 TCCGAGAGGGGCGGAGCGACACCCAGCCGCCCCGGGGCCCCCGGAACCCAGCGGGTG 372

QY 63 ATGGCAAAAGACCCGCTGTTGCGGAGCGGCTGCGAGAGCGCGGCGGTTAC 122
 DB 373 GCGACTAAGAGCCCGCGGCGCGGAGGAGCAACCGCGGAGAAATCGCCGAG 432
 QY 123 AACCGGACCCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
 DB 433 CCGAATCCGCGGCACTCCAGACGCGCGCGTGAAGCGCGCCAACTCCAGAGA 492
 QY 183 GGGTTCCT 242
 DB 493 CCGGCGCAGGCGCTGCGAGAAAGCTTGACCTTGAAGCGCGCCCAAAACCGGAG 552
 QY 243 CTCGAACAGCGCGCTGCGAGCGGAGTCCCTCGGAGTGAATGAGTCCAGCTCTCT 302
 DB 553 CCATGGAACCCCGCGGCTGCGGCTTAAAGAGCGGCTCTCTCTCTCTCTCTCTCT 612
 QY 303 CTCGCTCTCC 311
 DB 613 CTGGCGGCC 621

RESULT 9

AA26227

ID AAX26227 standard; DNA, 912 BP.

AC AAX26227;

XX 25-MAY-1999 (first entry)

DE HIV-1 VP22 polypeptide encoding DNA.

XX Cyclin-dependent kinase; CDK; CDK/cyclin complex; inhibitory; restenosis;
 KW CDK-binding motif; endothelialisation; fusion protein; therapeutic; acne;
 KW intracellular; transcellular; transcytosis; vascular wound; repair; hair;
 KW smooth muscle; cardiovascular; arteriosclerotic; fibrotic disorder;
 KW cellular proliferation; rheumatoid arthritis; diabetes; cirrhosis; graft;
 KW tumour; inflammation; neurodegeneration; periodontal; spermatogenesis;
 KW tachycardia; HIV-1; ss.

OS Human immunodeficiency virus type 1.

XX WO9906540-A2.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-US15759.

XX 29-JUL-1997; 97US-0902572.

XX (MITO-) MITOTIX INC.

XX Beach DH, Gyuris J, Lamphere U;

XX WPI; 1999-153770/13.

XX P-PSDB; AAM95099.

XX Fusion and chimeric proteins including cyclin-dependent kinase
 PT binding motif - used for regulation of cell proliferation and
 PT differentiation, for treatment of, e.g. vascular injury, cancers,
 PT fibrosis and neurodegeneration

XX Example 2; Page 26; 88pp; English.

XX The invention relates to novel inhibitors of cyclin-dependent kinases
 CC (CDKs), particularly CDK/cyclin complexes. It provides a recombinant
 CC transfection system (A) that comprises: (i) first gene construct
 CC comprising a sequence encoding an inhibitory polypeptide containing at
 CC least one CDK-binding motif for binding and inhibiting activity of a
 CC CDK, linked to a transcription regulator functional in eukaryotic cells;
 CC (ii) second gene construct comprising a sequence encoding a polypeptide
 CC that promotes endothelialisation, and (iii) a gene delivery composition
 CC for delivering the GEs to a cell for transfection. Also provided are

CC nucleic acids encoding a fusion protein (FP) containing: (i) a
 CC therapeutic polypeptide sequence (TP) from an intracellular protein that
 CC alters a cellular process when FP enters the cell, and (ii) a
 CC transcellular polypeptide sequence (TCP) that promotes transcytosis of
 CC FP. The FP consists of at least one CDK-binding motif and a TCP. See
 CC AAX26227 for detailed uses of the recombinant transfection system. The
 CC CKI polypeptides are engineered to include any of the peptides shown in
 CC AAM95097-100 encoded by the DNA sequences AAX26225-228.

SQ Sequence 912 BP; 149 A; 354 C; 287 G; 122 T; 0 other;

Query Match 2.8%; Score 48.2; DB 20; Length 912;

Best Local Similarity 47.2%; Pred. No. 0.0051;

Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTGAGTTTGAATCCGAGCCAGTCTCTCTCGGAGCCAGCGGTCCGAGCGG 62
 DB 325 TCCGAGGCGCGGAGCGAGCAACCCAGACGCGCCCGGCGCCCGGAAACCGAGCGGTG 384
 QY 63 ATGGCAAAAGACCCGCTGTTGCGGAGCGGCTGCGAGAGCGCGGCGGTTAC 122
 DB 385 GCGACTAAGAGCCCGCGGCGCGGAGGAGCAACCGCGGAGAAATCGCCGAG 444
 QY 123 AACCGGACCCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
 DB 445 CCGAATCCGCGGCACTCCAGACGCGCGCGTGAAGCGCGCCAACTCCAGAGA 504
 QY 183 GGGTTCCT 242
 DB 505 CCGGCGCAGGCGCTGCGAGAAAGCTTGACCTTGAAGCGCGCCCAAAACCGGAG 564
 QY 243 CTCGAACAGCGCGCTGCGAGCGGAGTCCCTCGGAGTGAATGAGTCCAGCTCTCT 302
 DB 565 CCATGGAACCCCGCGGCTGCGGCTTAAAGAGCGGCTTGAAGCGCGCGGTGCGGCGC 624
 QY 303 CTCGCTCTCC 311
 DB 625 CTGGCGGCC 633

RESULT 10

AA288468

ID AAZ88468 standard; DNA, 912 BP.

XX AAZ88468;

XX 10-MAY-2000 (first entry)

DE HSV-1 VP22 peptide encoding nucleotide sequence.

KW Papillomavirus; PV; infection; cell proliferation; E2; peptidomimetic;
 KW E1; antiviral; virucide; cytostatic; antiproliferative; dermatological;
 KW preneoplastic lesion; neoplastic lesion; cutaneous lesion; wart;
 KW epidermodysplasia verruciformis; anorectal carcinoma; ss.

OS Herpes simplex virus type 1.

XX WO200001720-A2.

XX 13-JAN-2000.

XX 02-JUL-1999; 99WO-US15144.

XX 02-JUL-1998; 98US-0091661.

XX (HARD) HARVARD COLLEGE.

XX Howley P, Benson J, Kasukawa H;

XX WPI; 2000-171001/15.

XX P-PSDB; AAY98877.

PT Use of papillomavirus E2 protein peptidomimetics for treating

KM Ubiquitin dependent proteolysis modulation; cdc4 phospho design motif;
 KM CDP motif; cytosolic; neotrophic; antiproliferative; cell proliferation;
 KM growth; differentiation; cancer; neurodegenerative disorder;
 KM spinal degeneration; gene; ds.

OS Herpes simplex virus.

Key Location/Qualifiers
 CDS 4..906
 /*tag= a

FT /partial
 FT /product= "HSV-1 VP22 protein"
 FT /transl_except= (pos:376..378, Gly)
 FT /note= "no stop codon given"

XX WO200183518-A2.

PD 08-NOV-2001.

PF 04-MAY-2001; 2001WO-CA00632.

XX 04-MAY-2000; 2000US-202166P.

PR 24-JAN-2001; 2001US-263774P.

PA (MOUNT) MOUNT SINAI HOSPITAL.

PI Nash P, Pawson T, Tang X, Tyers M;

DR WPI; 2002-164074/21.

XX P-PSDB; ABB05524.

PT New Cdc4 Phospho Design motif that targets molecules for ubiquitin
 PT dependent proteolysis, is useful for the modulation of cell
 PT proliferation i.e. cancer treatment -

XX Disclosure; Page 30; 83pp; English.

XX The present invention describes a cdc4 phospho design (CPD) motif, (C),
 CC that targets molecules for ubiquitin dependent proteolysis. (C) have
 CC cytosolic, neotrophic and antiproliferative activity. Also described is
 CC a method for the treatment of a disease or condition where affected
 CC cells have a defective protein, comprising administering (C) to promote
 CC degradation of the target protein in cells by ubiquitin dependent
 CC proteolysis. (C) can also be used for modulating the proliferation,
 CC growth and/or differentiation of cells. (C) can be used to modulate
 CC ubiquitin dependent proteolysis or cell proliferation, growth and/or
 CC differentiation of cells. (C) is useful in the treatment of cancers and
 CC neurodegenerative disorders as well as spinal degeneration. The present
 CC sequence encodes the HSV-1 VP22 protein which is given in the
 CC exemplification of the present invention.

XX Sequence 912 BP; 149 A; 354 C; 287 G; 122 T; 0 other;

Query Match 2.8%; Score 48.2; DB 24; Length 912;

Best Local Similarity 47.2%; Pred. No. 0.0051;

Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTGAGTTGATCCGAGCCGACAGTCTCTCTCGGCGCCACCGCGCTCCGACGGCG 62

DB 325 TCCGAGAGGCGCGACGACACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384

QY 63 ATGCAAGAGCCCGCTGCTTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122

DB 385 GCGACTAAGGCG 444

QY 123 AACCGAACCGAGCT 182

DB 445 CCAAGATTCG 504

QY 183 GGGTTCCT 242

DB 505 CCGCGAGAGGCG 564

QY 243 CTCACACCG 302

DB 565 CCAATGAGACCG 624

QY 303 CTCGCTCTCC 311

DB 625 CTGCG 633

RESULT 13

AAV17085

ID AAV17085 standard; DNA; 950 BP.

AC AAV17085;

DT 03-JUL-1998 (first entry)

DE Herpes simplex virus UL49 gene.

XX Herpes simplex virus.

OS Herpes simplex virus.

XX Key Location/Qualifiers

FT CDS 45..950

FT /*tag= a

FT /product= VP22

XX WO9804708-A1.

XX 05-FEB-1998.

XX 28-JUL-1997; 97WO-GB02036.

XX 26-JUL-1996; 96GB-0015726.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Hope RG, McGeoch DJ, McLaughlan J, Rixon HMM,

XX WPI; 1998-130696/12.

XX P-PSDB; AAW47194.

XX New antiviral agent disrupting binding of VP22 to VP16 or gB -

XX useful for treating infections caused by herpes simplex, e.g. cold

XX sores and chicken-pox

XX Example; Pages 49-50; 75pp; English.

XX The present sequence encodes the herpes simplex virus (HSV)

XX tegument protein VP22. VP22 was used in the preparation of a novel

XX antiviral agent, which inhibits the maturation and/or replication

XX of HSV by disrupting association between VP22 and VP16 and/or gB.

XX The agent can be used to treat, e.g. cold sores, genital herpes,

XX chickenpox and shingles.

XX Sequence 950 BP; 154 A; 365 C; 301 G; 130 T; 0 other;

Query Match 2.8%; Score 48.2; DB 19; Length 950;

Best Local Similarity 47.2%; Pred. No. 0.0052;

Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTGAGTTGATCCGAGCCGACAGTCTCTCTCGGCGCCACCGCGCTCCGACGGCG 62

DB 366 TCCGAGAGGCGCGACGACACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425

QY 63 ATGCAAGAGCCCGCTGCTTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122

DB 426 GCGACTAAGGCG 485

QY 123 AACCGAACCGAGCT 182

PT New nucleic acids encoding fusion polypeptide comprising intercellular
PT transport polypeptide linked to antigenic polypeptide, useful as
PT therapeutic vaccine for cancer and major chronic viral infections -
XX
PS Disclosure, Fig 7; 102pp; English.

CC The present invention relates to a new nucleic acid molecule that
CC encodes a fusion polypeptide. The fusion protein comprises a first
CC polypeptide comprising at least one intercellular transport polypeptide
CC and a second polypeptide comprising at least one antigenic polypeptide
CC or peptide. The invention also describes an optional linker peptide
CC linking the first and second polypeptide. The nucleic acid is useful as
CC a vaccine for enhancing immune responses, primarily cytotoxic T
CC lymphocyte responses to specific antigens such as tumour or viral
CC antigens. The compositions comprising the nucleic acids are especially
CC useful as a therapeutic vaccine for cancer and for major chronic viral
CC infections, as well as in the treatment of veterinary herpesvirus
CC infections, including equine or bovine herpesvirus, Marek's disease virus
CC in chickens and other fowls, animal retroviral diseases, pseudorabies
CC and rabies. The present nucleic acid sequence encodes the pCDNA3-VP22/E7
CC fusion protein of the invention.
XX

Sequence 1257 BP; 246 A; 429 C; 368 G; 214 T; 0 other;

Query Match 2.8%; Score 48.2; DB 24; Length 1257;

Best Local Similarity 47.2%; Pred. No. 0.0061;

Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTGAGTTTGAATCGAGCCGACAGTCTCTCTCGGGCCACCGCGTCCGACCGCG 62
DB 322 TCCTGAGAGGCGCGGACGACACCAACCGCCCGCGCCCGGACCGACCGCGGATG 381
QY 63 ATGGCAAAAGACCCCGTCTGCGCGTGGCGCGCGTGGCGGAGCCCGCGCGGTTAC 122
DB 382 GCGCTTAAGGCCCGCGCGCGCGCGCGGAGACCAACCGCGGAGAAATCGGCCAG 441
QY 123 AACCGAGCCAGTCT 182
DB 442 CCAGAAATCGCGCGCATCTCCAGAGCGCCCGCGTGGAGCGCGCCAGATCCAGACA 501
QY 183 GGGTTCT 242
DB 502 CCGCGGAGGCGCTGGCCAGAAACCTTGAACCGCGCCCGCAAAACCCCGACGCG 561
QY 243 CTCGAACGCGCGTGGCGGAGTCCCTCGGTTTATGAGTCCAGGCTCGTGTCTC 302
DB 562 CCAATGACCCCGCGGATGGCGGTTTAACAAGCGCGTCTTCTGCGCCGCGGTGGGCGC 621
QY 303 CTGCTCTCC 311
DB 622 CTGGCGGCGC 630

Search completed: July 4, 2003, 20:47:02
Job time : 292 secs

APPLICANT: Robbins, Alan K.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
FILE REFERENCE: From Multiple Transfected Epilomes
CURRENT APPLICATION NUMBER: US/09/130,114
CURRENT FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 2
LENGTH: 1931
TYPE: DNA
ORGANISM: EBNA
US-09-130-114-2

Query Match 2.9%; Score 49.2; DB 2; Length 1931;
Best Local Similarity 50.9%; Pred. No. 0.008;
Matches 117; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 97 TCGCCGAGGCGCGGCGGCGGTTCAACACGACCTCTCTCTCTCTCTCTG 156
DB 411 TCCCGTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCG 470
QY 157 CGGTGAGGCTCCGATCCACAGAGGTTCTCTCTCTCTCTCTCTCTCTCT 216
DB 471 TCT 530
QY 217 GCGACGCGCGTGGGACCGCGCGCTCAACACGCGGTGCGAGCGGAGTCCCTCG 276
DB 531 TCT 590
QY 277 GATTGATGAGTCAAGCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 326
DB 591 TCGTCT 640

RESULT 3
US-09-347-504-11
Sequence 11, Application US/09347504
Patent No. 6399075
GENERAL INFORMATION:
APPLICANT: Howley, Peter M.
APPLICANT: Benson, John
APPLICANT: Kasukawa, Hiroaki
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
FILE REFERENCE: HMV-041.01
CURRENT APPLICATION NUMBER: US/09/347,504
CURRENT FILING DATE: 1999-07-02
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 11
LENGTH: 912
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Ndel-EcoRI
OTHER INFORMATION: Fragment
US-09-347-504-11

Query Match 2.8%; Score 48.2; DB 4; Length 912;
Best Local Similarity 47.2%; Pred. No. 0.001;
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTCGAGTTGATCCGAGCCACAGTCTCTCTCTCTCTCTCTCTCTCT 62
DB 325 TCCGAGGCGCGGAGCGACACCCACCCGCGCGCGCGCGCGCGCGCGG 384
QY 63 ATGGCAAAAGACCCCTGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 385 GCGACTTAAGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 444
QY 123 AACCGAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182

DB 445 CCAAGATCCGCGCGCACTCCAGAGCCCGCGGTCGAGGCGCAACCCGATCCAGACA 504
QY 133 GGGTCTCTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
DB 505 CCGCGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 564
QY 243 CTCAGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302
DB 565 CAGTGAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 624
QY 303 CTGCTCTCC 311
DB 625 CTGCGCGCC 633

RESULT 4
US-09-230-421-1
Sequence 1, Application US/09230421
Patent No. 6200577
GENERAL INFORMATION:
APPLICANT: Medical Research Council
TITLE OF INVENTION: ANTI-HERPESVIRAL AGENTS AND ASSAYS
FILE REFERENCE: P18189C
CURRENT APPLICATION NUMBER: US/09/230,421
CURRENT FILING DATE: 1999-01-25
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 1
LENGTH: 950
TYPE: DNA
ORGANISM: HERPESVIRUS TYPE 1
US-09-230-421-1

Query Match 2.8%; Score 48.2; DB 4; Length 950;
Best Local Similarity 47.2%; Pred. No. 0.001;
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTCGAGTTGATCCGAGCCACAGTCTCTCTCTCTCTCTCTCTCTCT 62
DB 366 TCCGAGGCGCGGAGCGACACACCCCGCGCGCGCGCGCGCGCGG 425
QY 63 ATGGCAAAAGACCCCTGCTTGGGCGGCGGCGGCGGCGGCGGCGG 122
DB 426 GCGACTTAAGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 485
QY 123 AACCGAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
DB 486 CCAAGATCCGCGCGCACTCCAGAGCCCGCGGTCGAGCGGCGCAACCCGATCCAGACA 545
QY 183 GGGTCTCTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
DB 546 CCGCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 605
QY 243 CTCAGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302
DB 606 CAGTGAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 665
QY 303 CTGCTCTCC 311
DB 666 CTGCGCGCC 674

RESULT 5
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
NUMBER OF SEQUENCES: 52

RESULT 7

US-08-680-506-5
; Sequence 5, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-5

Query Match

2.7%; Score 46.6; DB 3; Length 954;
Best Local Similarity 53.6%; Pred. No. 0.003;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 79 CGTTCGGGCGGTGAGCGGCGGTGCGGAGGCGCGCGCGGTTCAACACCGGACCGACTCC 138
DB 125 CGCTGTCCCTCGGAGCGCACG 184
OY 139 TCCT 198
DB 185 AGCTCCGCGCGCGGTGCTGCTGCGCACGCGCACCGCTACCCACGCGCGCGGGAATCCGCGCC 244
OY 199 GTGCGCTTCGAGACCTTTCGACGCGCGTGTGGGAAACCCCGCGCGCTCTCAACACCGCGCTCG 258
DB 245 TGGCCGAGCTGCACGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 304
OY 259 C 259
DB 305 C 305

RESULT 8

US-08-680-506-8
; Sequence 8, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-8

Query Match

2.7%; Score 46.6; DB 3; Length 1347;
Best Local Similarity 53.6%; Pred. No. 0.0035;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 79 CGTTCGGGCGGTGAGCGGCGGTGCGGAGGCGCGCGCGGTTCAACACCGGACCGACTCC 138
DB 125 CGCTGTCCCTCGGAGCGCACG 184
OY 139 TCCT 198
DB 185 AGCTCCGCGCGGTGCTGCTGCGCACGCGCACCGCTACCCACGCGCGCGGGAATCCGCGCC 244

OY

199 GTGCGCTTCGAGACCTTTCGACGCGCGTGTGGGACCCCGCGCGGTTCAACACCGGACCGACTCC 258
DB 245 TGGCCGAGCTGCACGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 304

OY

DB

259 C 259
305 C 305

RESULT 9

US-08-680-506-6
; Sequence 6, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-6

Query Match

2.7%; Score 46.6; DB 3; Length 1587;
Best Local Similarity 53.6%; Pred. No. 0.0038;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 79 CGTTCGGGCGGTGAGCGGCGGTGCGGAGGCGCGCGCGGTTCAACACCGGACCGACTCC 138
DB 205 CGCTGTCCCTCGGAGCGCACG 284
OY 139 TCCT 198
DB 265 AGCTCCGCGCGGTGCTGCTGCGCACGCGCACCGCTACCCACGCGCGCGGGAATCCGCGCC 324
OY 199 GTGCGCTTCGAGACCTTTCGACGCGCGTGTGGGAAACCCCGCGCGCTCTCAACACCGCGCTCG 258
DB 325 TGGCCGAGCTGCACGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384
OY 259 C 259
DB 385 C 385

RESULT 10

US-08-680-506-4
; Sequence 4, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2233
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-4

Query Match

2.7%; Score 46.6; DB 3; Length 2233;
Best Local Similarity 53.6%; Pred. No. 0.0045;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 21:37:15 ; Search time 211 Seconds
(without alignments)

12591.504 Million cell updates/sec

Title: US-09-938-294-2

Perfect score: 1702
Sequence: 1 atccggcgcagattgacc.....gagagattcaattagc 1702

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262	15.4	276	10 US-09-923-876-6035	Sequence 6035, Ap
2	257	15.1	257	9 US-09-923-876-6078	Sequence 6078, Ap
3	52.4	3.1	2987	9 US-10-128-714-250	Sequence 250, App
4	52.4	3.1	3495	9 US-10-128-714-5250	Sequence 5250, App
5	49.8	2.9	1326	9 US-10-156-761-3022	Sequence 3022, Ap
6	49.8	2.9	9025608	9 US-10-156-761-1	Sequence 1, Appl1
7	48.2	2.8	912	8 US-08-902-572-13	Sequence 13, Appl1
8	48.2	2.8	912	9 US-10-161-499-11	Sequence 11, Appl1
9	48.2	2.8	912	10 US-09-862-179A-40	Sequence 40, Appl1
10	48	2.8	918	9 US-10-156-761-5655	Sequence 5655, Ap
11	48	2.8	9025608	9 US-10-156-761-1	Sequence 1, Appl1
12	47	2.8	520	9 US-10-184-644-332	Sequence 332, App
13	47	2.8	520	9 US-10-184-634-332	Sequence 332, App
14	46.2	2.7	1038	9 US-10-156-761-1894	Sequence 1894, Ap
15	46	2.7	1905	9 US-09-990-385-2	Sequence 2, Appl1
16	45.6	2.7	594	9 US-10-123-155-10	Sequence 10, Appl1
17	45.4	2.7	1431	9 US-10-128-714-7250	Sequence 7250, Ap
18	45.4	2.7	1495	9 US-10-128-714-6250	Sequence 6250, Ap
19	45	2.6	621	9 US-10-156-761-4870	Sequence 4870, Ap

20	45	2.6	978	9 US-10-156-761-1779	Sequence 1779, Ap
21	44.8	2.6	1662	9 US-10-156-761-3483	Sequence 3483, Ap
22	44.6	2.6	738	9 US-10-156-761-276	Sequence 276, App
23	44.4	2.6	2480	9 US-10-098-841-184	Sequence 184, App
24	44.4	2.6	2480	9 US-10-098-841-186	Sequence 186, App
25	44.4	2.6	2531	9 US-10-098-841-185	Sequence 185, App
26	44.2	2.6	2855	9 US-09-904-420A-1	Sequence 1, Appl1
27	43.8	2.6	41936	10 US-09-967-768A-116	Sequence 116, App
28	43.2	2.5	1143	9 US-10-156-761-5707	Sequence 5707, App
29	43	2.5	849	9 US-10-156-761-7075	Sequence 7075, App
30	43	2.5	1140	9 US-10-156-761-6044	Sequence 6044, App
31	42.8	2.5	1440	9 US-10-156-761-6190	Sequence 6190, App
32	42.8	2.5	2334	9 US-10-156-761-2734	Sequence 2734, App
33	42.8	2.5	2790	9 US-10-300-834-5	Sequence 5, Appl1
34	42.8	2.5	3432	9 US-10-300-834-4	Sequence 4, Appl1
35	42.6	2.5	1428	9 US-10-156-761-5428	Sequence 5428, App
36	42.4	2.5	390	10 US-09-960-352-14023	Sequence 14023, A
37	42.4	2.5	395	10 US-09-960-352-6638	Sequence 6638, App
38	42.2	2.5	4257	9 US-09-825-288A-1	Sequence 1, Appl1
39	42	2.5	5025	10 US-09-960-253-176	Sequence 176, App
40	42	2.5	42939	9 US-09-799-462A-17	Sequence 17, Appl1
41	42	2.5	42939	9 US-10-125-767-17	Sequence 17, Appl1
42	42	2.5	42939	9 US-09-836-911A-17	Sequence 17, Appl1
43	42	2.5	42939	9 US-10-151-081-17	Sequence 17, Appl1
44	42	2.5	42939	9 US-10-287-313-17	Sequence 17, Appl1
45	42	2.5	42939	9 US-10-219-694-17	Sequence 17, Appl1

ALIGNMENTS

RESULT 1
US-09-923-876-6035
Sequence 6035, Application US/09923876
Patent No. US20020013958A1
GENERAL INFORMATION:
APPLICANT: Lalugudi, Raghunath V.
APPLICANT: Kamigaki, Laura Y. (Ito)
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNICTOFORTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: PL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923, 876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/298,329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO 6035
LENGTH: 276
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No. US20020013958A1 700458234H1
NAME/KEY: unsure
LOCATION: 200, 275
OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-6035

Query Match 15.4%; Score 262; DB 10; Length 276;
Best Local Similarity 99.3%; Pred. No. 1.1e-72;
Matches 273; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy	1034	CAATTGGACCTCGCTTACCTGCTTGTGTGTAAGAACACGATTCAGACCGGTGCA	1093
Db	1	CAATTGGACCTCGCTTACCTGCTTGTGTGTAAGAACACGATTCAGACCGGTGCA	60
Qy	1094	TTTGTGACACAGACATTTGGACATGGGCCCTTACTTGGACAGCAATTTGATGCTTTGCA	1153
Db	61	TTTGTGACACAGACATTTGGACATGGGCCCTTACTTGGACAGCAATTTGATGCTTTGCA	120

QY 1154 GAATCTCAGGGCGGTGGAGATGCTTTGGAAGATTAACATTGAAGCAATGGCATTCAA 1213
DB 121 GAATCTCAGGGCGGTGGAGATGCTTTGGAAGATTAACATTGAAGCAATGGCATTCAA 180
QY 1214 GTTGCAGATTTGGGCAAGGCTGTGGAGGACACAGAGATGCTCTGGAAGGCTCGAC 1273
DB 181 GTTGCAGATTTGGGCAAGGCTGTGGAGGACACAGAGATGCTCTGGAAGGCTCGAC 240
QY 1274 TGGCCTTCTGCATCTCTGCTGGGAAGGAGGGCGTGG 1308
DB 241 TGGCC-TCTGCATCTCTGCTGGGAAGGAGGGCGTGG 274

RESULT 2
US-09-923-876-6078
Sequence 6078, Application US/09923876
Patent No. US2002013958A1
GENERAL INFORMATION:
APPLICANT: Laijudi, Raghunath V.
APPLICANT: Kamigaki, Laura Y. (Ico)
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: PL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923,876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/298,329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO 6078
LENGTH: 257
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inceye ID No. US20020013958A1 700458322H1
US-09-923-876-6078

Query Match 15.1%; Score 257; DB 10; Length 257;
Best Local Similarity 100.0%; Pred. No. 4e-71;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1035 AAATTGAGACCTGACCTGACCTTTGTGTGGAAGAACCAATTCATGACCGGTGCAT 1094
DB 1 AAATTGAGACCTGACCTGACCTTTGTGTGGAAGAACCAATTCATGACCGGTGCAT 60
QY 1095 TTGTGAAACAGACATTTGGAGAGTGCCCTTACTTGGCAGCAATTGATGCTTTGTCAG 1154
DB 61 TTGTGAAACAGACATTTGGAGAGTGCCCTTACTTGGCAGCAATTGATGCTTTGTCAG 120
QY 1155 AATTCTCAGGGCGGTGGAGATGCTTTGGAAGATTAACATTGAAGCAATGGCATTCAAG 1214
DB 121 AATTCTCAGGGCGGTGGAGATGCTTTGGAAGATTAACATTGAAGCAATGGCATTCAAG 180
QY 1215 TTGCAGATTTGGGCAAGGCTGTGGAGGACACAGAGATGCTCTGGAAGGCTCGAC 1274
DB 181 TTGCAGATTTGGGCAAGGCTGTGGAGGACACAGAGATGCTCTGGAAGGCTCGAC 240
QY 1275 GGCCTTCTGCATCTCTGCTGGGAAGGAGGGCGTGG 1291
DB 241 GGCCTTCTGCATCTCTGCTGGGAAGGAGGGCGTGG 257

RESULT 3
US-10-128-714-250
Sequence 250, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Weng
APPLICANT: Tishkoff, Daniel

APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 250
LENGTH: 2987
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-250

Query Match 3.1%; Score 52.4; DB 9; Length 2987;
Best Local Similarity 48.1%; Pred. No. 3.2e-05;
Matches 149; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 1173 GAATCTTTGGAAGATTAACATTGAAGCAATGGCATTCAAGTTCCAGATTTGGGACAG 1232
DB 2223 GAACACTTTGGTATTTGTCTCCGTGGAGCCATCGCGCGGGGATCCCGTCTCGATCA 2282
QY 1233 GCTGTGAGGAGACACAGAGATGCTCTGGAAGGCTGAGTGGCTTCTGCATCTCTGCT 1292
DB 2283 AATACAGGCGGCGCTTAAGAACTATGTTGAAGGCGAAACGGGCTGCTCGGAGCGCA 2342
QY 1293 GGAAGAGAGGCGGTGGCGGCTCTTGGCAAGAACATGTCAGACTGCAAGCCAGCCGAG 1352
DB 2343 ACGGTGTGCGGATTTGATGACCTGCGGTCATGACCGGGTCTCTAGATGAGCCAGAA 2402
QY 1353 CAGAGGCTCTCATTGGGGGAAAGGCTATGSCAGGCTGGAAGAAATGTTTCATGAGCAC 1412
DB 2403 GACTTGAACCGCATGACCGCTGTAGCAAGGCGAGGTGAGAAAGATTTCTGCTGACT 2462
QY 1413 CACATGCTGAGAGATGCGCGCGGTGTTGAAGATGCTCTGAGAAATCAGAGAGCAC 1472
DB 2463 GCATGGGCTGACAGGCTCGAGGAGAGATCACCATATGCTCAGCAGAAAGAGCGCG 2522
QY 1473 TCCAGGTCTT 1482
DB 2523 TTCAGCGGTT 2532

RESULT 4
US-10-128-714-5250
Sequence 5250, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Weng
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23

```

; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5250
; LENGTH: 3495
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5250

```

```

Query Match      3.1%; Score 52.4; DB 9; Length 3495;
Best Local Similarity 48.1%; Pred. No. 3.6e-05;
Matches 149; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```

```

QY 1173 GAATGCTTGAAGATTAACAATTAAGCAATGCAATTCGACATTTGGGACACG 1232
DB 2223 GAAACATTGTGATTTGTCCTCCGTGAAGCCATGCGCGGAGATCCCGTCTCCATCA 2282
QY 1233 GCTCTGAGAGGACACGAGATGCTCTGACGAGCTGACTGCGCTTCTGATCTGCT 1292
DB 2283 AATACAGCGCGGCAATTAAGAACTATCTGTGAAGGCGAAACGSGCTGCTCCGGACCA 2342
QY 1293 GGGAGAGAGGCGTGGCGCTCTTGAAGAAACATCTGACACCTCGACGACGCGGAG 1352
DB 2343 AGGTTGTCGGATGATGATCGCGTCTATGACCGGGTTCTCTACGATGACACGAG 2402
QY 1353 CAGAGGCTTCATCGAGGAGAAAGGCTATGACAGGCTGAAGAAATGTTTCATGAGCAC 1412
DB 2403 GACTTGACCGCATGTCAGCGCTGACAGAGGAGGAGAGGAGGAGTCTCGCTGACT 2462
QY 1413 CACATGCTGAGAGATCGCGCGTGTGAAGATGCTCTGAGAAATCAACAGGACAC 1472
DB 2463 GCAATGGGTACAGGCTGAGAGAGAGATACCGATATCTCACAAGAAAGACGCGCG 2522
QY 1473 TTCAGGCTTT 1482
DB 2523 TTCAGGCTTT 2532

```

```

RESULT 5
US-10-156-761-3022
; Sequence 3022, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3022
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1326)

```

```

US-10-156-761-3022

```

```

Query Match      2.9%; Score 49.8; DB 9; Length 1326;
Best Local Similarity 52.7%; Pred. No. 0.00012;
Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

```

QY 56 ACCGCGATGCGAAAGACCCCGTCTTGGCGGTGGCGGCGGTGCGGAGGCGCGGCGCC 115
DB 261 ACCGCGTTCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTGCGCTACGACTGCTCGGAGAA 320
QY 116 GATTCAACCGGACCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175
DB 321 CGCGCGGACCGGACCGCATACAGGCGCGCTCTCAAGTGTGCGCGTGTGCGCGCGCGCG 380
QY 176 CACAGAGAGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
DB 381 CCGTCTGCTGTGTGTGCTCCGACGTGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 440
QY 236 CGCGCGCTCAACACCGCGCGTCCG 260
DB 441 CGCGCTCTCAACCGCGGATTGCGC 465

```

```

RESULT 6
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (418715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

```

Query Match      2.9%; Score 49.8; DB 9; Length 9025608;
Best Local Similarity 52.7%; Pred. No. 0.079;
Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

```

QY 56 ACCGCGATGCGAAAGACCCCGTCTTGGCGGTGGCGGCGGTGCGGAGGCGCGGCGCC 115
DB 3794466 ACCGCGTTCGCGCGCGCTGCGCGCGCGCGCGCGCGCTGCGCTACGACTGCTCGGAGAA 3794407
QY 116 GATTCAACCGGACCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175
DB 3794406 CGCGCGGACCGGACCGCATACAGGCGCGCTCTCAAGTGTGCGCGTGTGCGCGCGCG 3794347
QY 176 CACAGAGAGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
DB 3794346 CCGTCTGCTGTGTGTGCTCCGACGTGACCGCGGCGCGCGCGCGCGCGCGCGCG 3794287
QY 236 CGCGCGCTCAACACCGCGCGTCCG 260
DB 3794286 CGCGCTCTCAACCGCGGATTGCGC 3794262

```

```

US-08-902-572-13
; Sequence 13, Application US/08902572
Patent No. US20020068706A1
GENERAL INFORMATION:
APPLICANT: Gyuris, Jeno
APPLICANT: Lamphere, Lou
APPLICANT: Beach, David H.
TITLE OF INVENTION: INHIBITORS OF CELL-CYCLE PROGRESSION AND
TITLE OF INVENTION: RELATED THERETO
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELLIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/902,572
FILING DATE: 29-JUL-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV-069, 03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA.
US-08-902-572-13

Query Match 2.88; Score 48.2; DB 8; Length 912;
Best Local Similarity 47.28; Pred. No. 0.0003;
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0

QY 3 TCGGCTCAGATTATTCGAGGCCACAGTCTCTCTCGGAGCCACAGCGTTCGACGGCG 62
DB TCGGAGGGGCGGAGCCAGACCCACCAACCGCCGCCCGGGGCGCCGAAACCGAGGGG 384

QY 63 ATGGCAAGAGACCCCGCTGTTCCGGGTGGCGGGCGGTGCGCGAGAGCCGCGGGCGGTTCA 122
DB GCACATAGAGCCCCCGGCGCCCGGCGGAGAGCACCACCGCGAGAAATGGCCAG 444

QY 123 AACGGAGCCAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
DB CAGAAATCGCGCGACATCCCAAGACCCTCCCGCTCGACGGCGCAACCCGATCCAAAC 504

QY 183 GAGTCTCTCTCTCGCGGTGCGCTGCGAGACCTTGGCGAGCGCGTGGAGACCCGCGGC 242
DB CCGGCGAGGGGCTGGCGAAGAGCTGACATTAGACCGCGCCCCCAACCCGAGCGG 564

QY 243 CTCACACCGCGCTGCGAGAGCGGAGAGTCCCTCGGGTTCAAGAGTCTCAAGCTGTGCTC 302
DB CATTGACCCCCCGGGTGGCGGCTTTAACAAGCGGTCTTCTGCGCGCGGTGGGCGC 624

QY 303 CTCGCTCC 311
DB 625 CTGGCGGCC 633

```

```

RESULT 8
US-10-161-499-11
: Sequence 11, Application US/10161499
: Publication No. US20030044427A1
GENERAL INFORMATION:
APPLICANT: Howley, Peter M.
APPLICANT: Benson, John
APPLICANT: Kasukawa, Hiroaki
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
: FILE REFERENCE: HMV-041.01
: PAPER/PATENT STATUS: PENDING
CURRENT APPLICATION NUMBER: US/10/161,499
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US/09/347,504
PRIOR FILING DATE: 1999-07-02
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 912
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Ndel-EcoRI
US-10-161-499-11

Query Match          2.8%, Score 48.2, DB 9, Length 912;
Best Local Similarity 47.2%, Pred. No. 0.0003;
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTCGAGATTGATCCGAGCCACAGCTCTCTCGCGCGCCGCCACCGCGTCCGACCGCG 62
Db 325 TCGGAGAGGGGCGGAGCGAGACCCACACCGCGCCGCCCGCGCGCGCGCGCGCGCGCG 384

QY 63 ATGCGCAAAACCCCGTCTGCTTCGCGGTGGCGGCGCGTCCCGCGAAGCCCGCGCGCGTTCAC 122
Db 385 GCGCATMAAGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444

QY 123 AACCGAGCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
Db 445 CCGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 504

QY 193 GGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 242
Db 505 CCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 564

QY 243 CTCGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 302
Db 565 CCGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 624

QY 303 CTGGTCTCC 311
Db 625 CTGGCGCGCC 633

RESULT 9
US-09-862-179A-40
: Sequence 40, Application US/09862179A
: Patent No. US20020147306A1
GENERAL INFORMATION:
APPLICANT: Lin, Danny
APPLICANT: Pearson, Anthony
: TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
: FILE REFERENCE: WTSI-P01-009
: CURRENT APPLICATION NUMBER: US/09/862,179A
: CURRENT FILING DATE: 2001-05-21
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn version 3.1
SEQ ID NO 40
LENGTH: 912
TYPE: DNA

```

ORGANISM: HSV-1
US-09-862-179A-40

Query Match 2.8%; Score 48.2; DB 10; Length 912;
Best Local Similarity 47.2%; Pred. No. 0.0003;
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 3 TCGGCTGAGTTGATCCGAGCCGACAGTCTCTCTCGAGCCGACCGCGCTCCGACCGCG 62
Db 325 TCCGAGGAGGCGGAGGACGACACCAACGCGGCGGCGGCGGCGGCGGCGGCGG 384
Qy 63 ATGCAAAAGACCCCT 122
Db 385 GCGACTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 444
Qy 123 AACGAGACCGAGCT 182
Db 445 CCAAGATCCG 504
Qy 183 GGGTTCT 242
Db 505 CCGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 564
Qy 243 CTCAACACCGCGCT 302
Db 565 CCAATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 624
Qy 303 CT 311
Db 625 CTGGGCGGCG 633

RESULT 10

US-10-156-761-5655
; Sequence 5655, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5655
; LENGTH: 918
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(918)
US-10-156-761-5655

Query Match 2.8%; Score 48; DB 9; Length 918;
Best Local Similarity 50.0%; Pred. No. 0.00034;
Matches 120; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 73 CCGCGTCT 132
Db 566 CCGGATCCCGGAGACCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 625
Qy 133 AGCT 192
Db 626 AAGCT 685

Qy 193 TCCGCGTCT 252
Db 686 CCGGCT 745
Qy 253 CCGTCT 312
Db 746 CCGGCT 805

RESULT 11

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 2.8%; Score 48; DB 9; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 0.3;
Matches 120; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 73 CCGCGTCT 132
Db 6860939 CCGGATCCCGGAGACCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6860998
Qy 133 AGCT 192
Db 6860999 AACCT 6861058
Qy 193 TCCGCGTCT 252
Db 6861059 CCGGCT 6861118
Qy 253 CCGTCT 312
Db 6861119 CCGGCT 6861178

RESULT 12

US-10-184-644-332/C
; Sequence 332, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jiam
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.

Query Match 2.8%; Score 47; DB 9; Length 520;
Best Local Similarity 18.6%; Pred. No. 0.00047;
Matches 80; Conservative 89; Mismatches 262; Indels 0; Gaps 0

RESULT 13
US-10-184-634-332/c

Sequence 324, Application US/10/07924
Publication No. US20030068684A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME

Query Match	2.8%;	Score 47;	DB 9;	Length 520;
Best Local Similarity	18.6%;	Pred. No. 0.00047;		
Matches 80; Conservative	89;	Mismatches 262;	Indels 0;	Gaps 0;

RESULT 14
US-10-156-761-1894

```

Sequence 1634, Application US/70156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: ISEIDA, HARUO
APPLICANT: ICHIKAWA, JUN
APPLICANT: HORIIKAWA, HIROSHI
APPLICANT: SHIBATA, TATSUYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASHIYUKI
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDE
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-06-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1694
LENGTH: 1038
TYPE: DNA
ORGANISM: Streptomyces avermitilis

```

ORGANISM: *Streptomyces avermitilis*

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 18:55:10 ; Search time 1706 Seconds
(without Alignments)
16157.514 Million cell updates/sec

Title: US-09-938-294-2

Perfect score: 1702

Sequence: 1 atcgcgcgcaggttgatcc.....gaggagattcaatcttagt 1702

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

listing first 45 summaries

1: em_estba:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hiv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	815.4	47.9	843 11	AY107383 Zea mays
2	615	36.1	615 11	AY108664 Zea mays
3	612	36.0	613 10	AM067129 683016A01
4	551	32.4	678 12	BC464581 EML 71.F0
5	519.6	30.5	721 12	BF623089 HVSMB000
6	512	30.1	670 14	BC788642 WHEA152_B

7	502.6	29.5	557 12	BC464257 EML 71.F0
8	492.2	28.9	497 12	BC268271 1000172A0
9	417	24.5	627 13	B1959909 HVSMB002
10	401.6	23.6	439 12	BS650599 RH122 59
11	395	23.2	685 14	BQ148371 NF06707F
12	375.2	22.0	641 10	B1271361 NF050A09F
13	360.4	21.2	643 10	AW186747 ENLGH1137
14	342.4	20.1	375 10	AW257867 HFO2101F
15	342.4	19.2	437 14	BQ464581 HFO2101F
16	325.4	19.1	551 13	B1421630 EST532296
17	314.4	18.5	513 10	BE402653 CSB010B11
18	314.4	18.5	513 14	B0608401 BRY 4306
19	307	18.0	543 14	B0296254 san89a12.
20	302.8	17.8	443 14	C74729 C74729 Rice
21	218.4	12.8	461 10	BE209946 sc6h10.Y
22	215.8	12.7	390 10	BE022039 sm65906.Y
23	211.2	12.4	396 13	BJ208273 BJ208273
24	203.2	11.9	563 12	BF270892 GA_Eb000
25	203.2	11.9	711 10	AM266508 L0-1363T3
26	202.6	11.9	782 13	B1971272 GM830013A
27	186.2	10.9	425 14	T76431 11209 Lambd
28	185	10.9	360 9	AJ473497 AJ473497
29	185	10.9	360 9	AJ473498 AJ473498
30	171.2	10.1	863 17	BH501249 BGVX70TF
31	160.4	9.4	715 17	B78162 T31117TF TA
32	158.2	9.3	553 10	BE472046 EST416899
33	154	9.0	634 13	BJ173349 BJ173349
34	150.2	8.8	251 13	B1121818 F047P71Y
35	148.6	8.7	647 12	BF642258 NF066A10I
36	135.8	8.0	276 10	AV540026 AV540026
37	134.2	7.9	629 17	B21743 T31117-T7 T
38	131.6	7.7	669 13	BJ171310 BJ171310
39	129.6	7.6	629 17	AU239738 AU239738
40	125.8	7.4	288 17	BH810193 SALK_0442
41	125.4	7.4	435 14	BO662962 HS04P17u
42	104.6	6.1	569 13	B1267795 NF111D09I
43	102.2	6.0	688 14	BQ114748 EST600324
44	101.6	6.0	707 14	B0118269 EST603845
45	101.6	6.0	714 14	BQ118358 EST603934

ALIGNMENTS

RESULT 1
AY107383 843 bp mRNA linear HTC 25-MAY-2002

LOCUS Zea mays PC0096967 mRNA sequence.

DEFINITION Zea mays AY107383

ACCESSION Zea mays AY107383.1 GI:21210461

VERSION Zea mays

KEYWORDS Zea mays

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 843)

AUTHORS Hatney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of

JOURNAL Overgo Probes

REFERENCE 2 (bases 1 to 843)

AUTHORS Coe,E.C.

TITLE Direct Submission

JOURNAL Submitted (25-Apr-2002) Maize Mapping Project, University of

FEATURES Location/Qualifiers

1..843 /organism="Zea mays"

/db_xref="MaizeDB:655540"

/db_xref="taxon:4577"

/clone="PC0096967"

/clone.lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT 184 a 231 c 224 g 199 t 5 others
ORIGIN

Query Match 47.9%; Score 815.4; DB 11; Length 843;
Best Local Similarity 99.2%; Pred. No. 5,5e-200;

Matches 827; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 19 CCGAGCCCAAGTCTCTCTCTGCGCCACCGGTCGACGCGCGCATGCGAAGACCCCGT 78
DB 10 CCGAGCCCAAGTCTCTCTCTGCGCCACCGGTCGACGCGCGCATGCGAAGACCCCGT 69
QY 79 CGTTCCGGTGG 138
DB 70 CGTTCCGGTGG 129
QY 139 TCCT 198
DB 130 TCCT 189
QY 199 GTGCGCTCGAGACCTTTGCGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 258
DB 190 GTGCGCTCGAGACCTTTGCGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 249
QY 259 CCGAGCGGAGTCCCTCGGGTTCAAGAGTCCAGCTCTGCTCTCTCTCTCTCTCTCTCT 318
DB 250 CCGAGCGGAGTCCCTCGGGTTCAAGAGTCCAGCTCTGCTCTCTCTCTCTCTCTCTCT 309
QY 319 TCCT 378
DB 310 TCCT 369
QY 379 GCTTCGAAAGTGTGCTTAACCAACAGCATGCAAGAACAAATGATGTCATATTA 438
DB 370 GCTTCGAAAGTGTGCTTAACCAACAGCATGCAAGAACAAATGATGTCATATTA 429
QY 439 GCTTCGAAAGTGTGCTTAACCAACAGCATGCAAGAACAAATGATGTCATATTA 498
DB 430 GCTTCGAAAGTGTGCTTAACCAACAGCATGCAAGAACAAATGATGTCATATTA 489
QY 499 CAGTTGATTTGCTTAAGAGTCTGTTATCTTAACATCTGCTGCTGCGAAGT 558
DB 490 CAGTTGATTTGCTTAAGAGTCTGTTATCTTAACATCTGCTGCTGCGAAGT 549
QY 559 GGGTTGACCTGTTCTGAAGATCATGTTCTTAAGTCTTCCGAAGATTTTGTGTGA 618
DB 550 GGGTTGACCTGTTCTGAAGATCATGTTCTTAAGTCTTCCGAAGATTTTGTGTGA 609
QY 619 TCCATGAATGCGTGGGCTTACTTTAAGTTGAATGTCACAAATCTTCCCTTGTG 678
DB 610 TCCATGAATGCGTGGGCTTACTTTAAGTTGAATGTCACAAATCTTCCCTTGTG 669
QY 679 CTGAGACCATGATTGTTCTCATACAAGCGCTAGATTGGAATGACAGATCAGCATC 738
DB 670 CTGAGACCATGATTGTTCTCATACAAGCGCTAGATTGGAATGACAGATCAGCATC 729
QY 739 GCGTGAATAACAGATGCAAACTTATGTTGTCACCTGGGGAATGCTAATAAATA 798
DB 730 GCGTGAATAACAGATGCAAACTTATGTTGTCACCTGGGGAATGCTAATAAATA 789
QY 799 TGAAGTTGCTGAAGACAAT-GTGGCAAGAGTCTTACGGGAACATATTCGT 851
DB 790 TGAAGTTGCTGAAGACAAGTGTGGCAAGAGTCTTACGGGAACATATTCGT 843

RESULT 2

AY108664
LOCUS AY108664 615 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays PC0096986 mRNA sequence.
ACCESSION AY108664
VERSION AY108664.1 GI:21211836
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS Hayley, C.E., Dolan, M., Miao, G.H., Vogel, J.M., Whitel, M.S.,
1 (bases 1 to 615)
2 (bases 1 to 615)
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE
AUTHORS Coe, E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES
source
1..615
/organism="Zea mays"
/db_xref="taxon:4577"
/clone="PC0096986"
/clone.lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT 174 a 123 c 164 g 154 t
ORIGIN

Query Match 36.1%; Score 615; DB 11; Length 615;
Best Local Similarity 100.0%; Pred. No. 3.6e-148;

Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 ATGAGAGGACATGATGCGCTGAAGAAATACAGATGCGCAAACTTATGTTGTCACCTGG 780
DB 1 ATGAGAGGACATGATGCGCTGAAGAAATACAGATGCGCAAACTTATGTTGTCACCTGG 60
QY 781 GGAATAGTAAGAACTAATGGAAGTGTCTGAAGACATGTCGAAGAGTCTTACGGG 840
DB 61 GGAATAGTAAGAACTAATGGAAGTGTCTGAAGACATGTCGAAGAGTCTTACGGG 120
QY 841 AACATATTGCGAATCCCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 900
DB 121 AACATATTGCGAATCCCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
QY 901 GTGATACGAGGAAGGGAACAAGCTTATTTCTTCAACATTTTACAGCTTTCAGC 960
DB 181 GTGATACGAGGAAGGGAACAAGCTTATTTCTTCAACATTTTACAGCTTTCAGC 240
QY 241 TCATCAACAGGAAACTTAAGGCTGAATACATGCTGATGTTGGGAAGATG 300
DB 1021 TTAATGCTCAGACCAATTTGAGTCACTTACGATCTTGTGTGAAGACAGATTC 1080
QY 301 TTAATGCTCAGACCAATTTGAGTCACTTACGATCTTGTGTGAAGACAGATTC 360
QY 1081 ATGACCGTGTCAATTTTGTGAACAAGACATTTGCAAGTGGCCCTTACCTGCAAGATTG 1140
DB 361 ATGACCGTGTCAATTTTGTGAACAAGACATTTGCAAGTGGCCCTTACCTGCAAGATTG 420
QY 1141 ATGTCCTTTTCAAGATTTCTGAGGCGGTGGAAGATGCTTTGAAGATTAACATTTGAAG 1200

Db 421 ATGTCTTGTTCAGATTCTCAGGGCCGTGGGAATGCTTGGAGGATTAACATTTGAG 480

Qy 1201 CAATGGCATTCAGTTGCGAGATTGGGCAAGGCTGTGGAGGAGCAAGAGATGCTCC 1260

Db 481 CAATGGCATTCAGTTGCGAGATTGGGCAAGGCTGTGGAGGAGCAAGAGATGCTCC 540

Qy 1261 TGAAGCGCTGACTGCTCTTCTGCAATCTCTGCTGGAGAGAGGCGCTCTTTGGA 1320

Db 541 TGAAGCGCTGACTGCTCTTCTGCAATCTCTGCTGGAGAGAGGCGCTCTTTGGA 600

Qy 1321 AGAATCATGTCAGAC 1335

Db 601 AGAATCATGTCAGAC 615

RESULT 3

AM067129/c 633 bp mRNA linear EST 12-OCT-1999

LOCUS 683016A01.x1 683 - 14 day immature embryo from Hake lab (HS) Zea

DEFINITION mays cDNA, mRNA sequence.

ACCESSION AM067129

VERSION AM067129.1 GI:6022201

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 633)

REFERENCE Maize ESTs from various cDNA libraries sequenced at Stanford

AUTHORS Walbot V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

COMMENT Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 683016 row: A column: 01.

Location/Qualifiers

1. 633

/organism="Zea mays"

/cultivar="973"

/db_xref="taxon:4577"

/clone_lib="683 - 14 day immature embryo from Hake lab (HS)"

FEATURES

source

1. 633

/organism="Zea mays"

/cultivar="973"

/db_xref="taxon:4577"

/clone_lib="683 - 14 day immature embryo from Hake lab (HS)"

/tissue_type="embryo"

/dev_stage="14 days after pollination"

/lab_host="DH10B"

/note="Organ: embryo; Vector: PBKCMV (Stratagene's Zap Express); Site 1: XhoI; Site 2: EcoRI; Directionally cloned, 14 day immature embryo library created with Stratagene's Zap Express cDNA protocol."

BASE COUNT 161 a 174 c 134 g 159 t 5 others

ORIGIN

Query Match 36.0%; Score 612; DB 10; Length 633;

Best Local Similarity 99.2%; Pred. No. 2.2e-147;

Matches 612; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1079 TCATGACCGTGCATTTGTGACACAGACATTCGAGGCGCTTACTTGCGAGCAT 1138

Qy 633 TCATGACCGTGCATTTGTGACACAGACATTCGAGGCGCTTACTTGCGAGCAT 574

Db 1139 TCATGCTGTTTCAGAAATTCACAGGCGCTGAGAAATGCTTTGAAAGATTAACATTTGA 1198

Qy 573 TCATGCTGTTTCAGAAATTCACAGGCGCTGAGAAATGCTTTGAAAGATTAACATTTGA 514

Db 1199 AGCATGGCATTCAGATTCGACGATTCGAGGCGCTGAGAGGAGCAAGAGATGCT 1258

Qy 513 AGCATGGCATTCAGATTCGACGATTCGAGGCGCTGAGAGGAGCAAGAGATGCT 454

Qy 1259 CTGAGCGGCTGACTGAGCTTCTGCTGCTGCTGCTGAGAGAGGCGCTTTCG 1318

Db 453 CTGAGCGGCTGACTGAGCTTCTGCTGCTGCTGAGAGAGGCGCTTTCG 394

Qy 1319 AAGAAATCTGCTGACTGCTGAGAAATTCAGAGGAGCTTTCAGAGGAGGAGG 1378

Db 393 AAGAAATCTGCTGACTGCTGAGAAATTCAGAGGAGCTTTCAGAGGAGGAGG 334

Qy 1379 CTATGCGAGGAGTGAAGAAATGTTCAATGAGAGACACATGCTGAGAGGAGT 1438

Db 333 CTATGCGAGGAGTGAAGAAATGTTCAATGAGAGACACATGCTGAGAGGAGT 274

Qy 1439 GTTGAAGAGATTCCTGAGAGAAATTCAGAGGAGCTTTCAGAGGAGGAGG 1498

Db 273 GTTGAAGAGATTCCTGAGAGAAATTCAGAGGAGCTTTCAGAGGAGGAGG 214

Qy 1499 ATCAGCTTCGCTAACATGTTGAATTTTACGAGGCTTTCAGAGGAGGAGT 1558

Db 213 ATCAGCTTCGCTAACATGTTGAATTTTACGAGGCTTTCAGAGGAGGAGT 154

Qy 1559 GTAACTGTAATTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1618

Db 153 GTAACTGTAATTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 94

Qy 1619 CCATGTCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1678

Db 93 CCATGTCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 34

Qy 1679 GATTGAGGAGATTCA 1695

Db 33 GATTGAGGAGATTCA 17

RESULT 4

BG464581 678 bp mRNA linear EST 20-MAR-2001

LOCUS BM1_71.E08.g1_A002 Embryo 1 (BM1) sorghum bicolor cDNA, mRNA

DEFINITION sequence.

ACCESSION BG464581

VERSION BG464581.1 GI:13393230

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 678)

REFERENCE Reid, S.P., Cordomier-Pratt, M.-M., Gingle, A. and Pratt, L.H.

Arid, S.P., Cordomier-Pratt, M.-M., Gingle, A. and Pratt, L.H.

An EST database from Sorghum: developing embryos

Unpublished (2000)

Contact: Cordomier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: POLYTMix

High quality sequence start: 26

High quality sequence stop: 669

POLYA=No.

Location/Qualifiers

1. 678

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Embryo 1 (BM1)"

/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from polyA RNA in the cloning

vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 186 a 135 c 191 g 164 t 2 others

Query Match 32.4%; Score 551; DB 12; Length 678;

Best Local Similarity 95.0%; Pred. No. 1.3e-111;

Matches 569; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Db 890 AATTAATTAACAGTGTATCAGAGAAAGGACAAAGACTTTTCTTCAAGCATTTTATCA 949

1 AATTAATTAACAGTGTATCAGAGAAAGGACAAAGACTTTTCTTCAAGCATTTTATCA 60

950 GGGTTTGCAGTCAATCCAGACAGAAATTTAAAGTCCCTAGAAATCAATGCTAGATTGT 1009

61 GGGTTTGCAGTCAATCCAGACAGAAATTTAAAGTCCCTAGAAATCAATGCTAGATTGT 120

1010 GGGAAATGATGTAAATGCTCAAGCAAAATTTGACATTCATTCAGTGTGTGTA 1069

121 GGGAAATGATGTAAATGCTCAAGCAAAATTTGACATTCATTCAGTGTGTGTA 180

1070 GAAACAGATTCATGACCGTGTCCATTTGTGAACAAGCATTTGGCAGTGGCCCTTACTT 1129

181 GAATGGATTCATGACCGTGTCCATTTGTGAACAAGCATTTGGCAGTGGCCCTTACTT 240

1130 GGGAGCAATGATGTGTCTGTTCAGATTTTCAGGGCCCTGAGAAATGCTTTGGAAGAT 1189

241 GGGAGCAATGATGTGTCTGTTCAGATTTTCAGGGCCCTGAGAAATGCTTTGGAAGAT 300

1190 AACCAATGGAAGCAATGAGCATTCAGATTTGGGAGCCGCTGCTGAGAGGACAC 1249

301 AACCAATGGAAGCAATGAGCATTCAGATTTGGGAGCCGCTGCTGAGAGGACAC 360

1250 GGAATGCGCTCGAGAGCGGCTCGACCTGCTTCGATCTCGTGGGAGAGAGGCGCTGGC 1309

361 GGAATGCGCTCGAGAGCGGCTCGACCTGCTTCGATCTCGTGGGAGAGAGGCGCTGGC 420

1310 GCGCTTTCGAAAGAAATCTGTCAAGTCCGAGCCGAGCAGAGAGGCTTCATGGG 1369

421 GCGCTTTCGAAAGAAATCTGTCAAGTCCGAGCCGAGCAGAGAGGCTTCATGGG 480

1370 GGAAGAGGCTATGCGAGGGGTGAAGAAATGTTCAATGAGCACCACATGCGTGAAGAT 1429

481 GGAAGAGGCTATGCGAGGGGTGAAGAAATGTTCAATGAGCACCACATGCGTGAAGAT 540

1430 GCGGCGGCTGTGAAGATGCTCTGAAGAAATCAGAGAGCACTCAAGTCTTGAAGCTT 1488

541 GCGGCGGCTGTGAAGATGCTCTGAAGAAATCAGAGAGCACTCAAGTCTTGAAGCTT 599

RESULT 5
LOCUS BFe29089 721 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMDB0009N03f Hordeum vulgare seedling shoot EST library
HVCNDA0002 (Dehydration stress) Hordeum vulgare cDNA clone
BFe29089
BFe29089.2 GI:13091072
EST.
Hordeum vulgare.
SOURCE
ORGANISM
Hordeum vulgare.
Eukaryota; Viridiplantae; Scrophophylla; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
Triticaceae; Hordeum.
1 (bases 1 to 721)
Wing, R., Close, T.J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu,
Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton,
R.D., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomes: Morex drought-stressed seedling shoot cDNA
library
Unpublished (2001)
On Dec 19, 2000 this sequence version replaced gi:11893247.

Contact: Wing R
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 588
Seq primer: AATTACCTCCTCAATAGG
High quality sequence stop: 716.
Location/Qualifiers
1. 721
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMDB0009N03f"
/clone_1b="Hordeum vulgare seedling shoot EST library
HVNDN0002 (Dehydration stress)"
/issue_type="seedling shoot"
/lab_host="TUC121"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, mycelatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 90% RH for 24 hr. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, 600000 pfu were
in vivo excised to give phagescript SK(-) cDNA phagemids.
These steps were performed in the TJ Close Laboratory at
the University of California, Riverside (Choi, Close,
Fenton). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI.
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinholz A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomes. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/g9p/pages/bgm/31/cover.html)"

BASE COUNT 195 a 151 c 203 g 172 t

Query Match 30.5%; Score 519.6; DB 12; Length 721;

Best Local Similarity 83.7%; Pred. No. 1.8e-123;

Matches 601; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

Db 781 GGAATGATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

1 GGAATGATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60

841 AACATATTCGATGATCCCTTGAAGTGAAGATGATGATGATGATGATGATGATGATGAT 900

61 AACATATTCGATGATCCCTTGAAGTGAAGATGATGATGATGATGATGATGATGATGAT 120

901 GGTGATCAGGAGAAAGGACAAATTTTCTTCAAGCATTTTATCAGGCTTTCAGG 960

121 GGTGATCAGGAGAAAGGACAAATTTTCTTCAAGCATTTTATCAGGCTTTCAGG 180

961 TATATCAACAGAGAAATTAAGGCTAGATATATATATATATATATATATATATATATAT 1020

181 TATATCAACAGAGAAATTAAGGCTAGATATATATATATATATATATATATATATATAT 240

1021 TTAATGCTCAAGCAATTTGAGATCAGTACGATCTTGTGTGAAGAACATTC 1080

241 TGAATGCTCAAGCAATTTGAGATCAGTACGATCTTGTGTGAAGAACATTC 300

QY 1081 ATGACCGTGTCCATTTTGTGAACAAGACATTTGGAGTGGCCCTTACTTGTGGAGCAATTG 1140
 Db 301 ATGACCGTGTTCATCTTTGTCAACAAGACGTTGGCAGTGGACCTTATTGGCTGGCAATTG 360
 QY 1141 ATGTGCTTTGTCAAGAAATTCACAGGCGCTGGAGAAATGCTTTGGAAAGATTAACAATTGAAG 1200
 Db 361 ATGTGCTTTGTCAAGAAATTCACAGGCGCTGGAGAAATGCTTTGGAAAGATTAACAATTGAAG 420
 QY 1201 CAATGGCATTCAAGTTGGCCAGTATTTGGGACCGGCTGTGTGAAGGACCAAGAGATGCTCC 1260
 Db 421 CAATGGCATTCAAGTTGGCCAGTATTTGGGACCGGCTGTGTGAAGGACCAAGAGATGCTCC 480
 QY 1261 TGGACGGCTCGACTGGCTCTTGTGATCTCTTGGAAAGAGAGGGCTTCTTGGAA 1320
 Db 481 TGGAGGTTCACACGGCTCTTGTGATCTCTTGGAAAGAGAGGGCTTCTTGGAA 540
 QY 1321 AGAACATGTCAGACTCGACAGCCAGCCAGAGAGGCTTCATCGAGGAGAAAGAGCT 1380
 Db 541 GGAACATGTCAGACTCGACAGCCAGCCAGAGAGGCTTCATCGAGGAGAAAGAGCT 600
 QY 1381 ATGGCAGGTTGAAGAAATGTTTCATGGAGACCAATGCTGAAGAGATCGCGGGGCT 1440
 Db 601 ATGGCAGGTTGAAGAAATGTTTCATGGAGACCAATGCTGAAGAGATCGCGGGGCT 660
 QY 1441 TGAAGATGTCCT--GAGAAATCAGAGAGCACTCCAGGTCCTTGAAGCTTGGCGTG 1495
 Db 661 TGAAGATGTCCTTTCACAAAGAGTCTCGAGAGATCTCATCTTGAAGCTTGCATG 718

RESULT 6
 LOCUS B0788642 670 bp mRNA linear EST 26-JUL-2002
 DEFINITION WHE4152_B09_C182S wheat CS whole plant cDNA library Triticum
 accession B0788642
 VERSION B0788642.1 GI:21997114
 KEYWORDS EST.
 SOURCE bread wheat
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticaceae; Triticum.
 1 (bases 1 to 670)
 Anderson, O.D., Rausch, E., Chao, S., Crossman, C., Deal, K., Dvorak, J.,
 Lazo, G.R., Khush, C.J., Wilson, C., and Woo, J.
 The structure and function of the expressed portion of the wheat
 genomes Chinese Spring whole plant cDNA library
 Unpublished (2002)
 JOURNAL Contact: Olin Anderson
 COMMENT US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@w.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer.
 Location/Qualifiers

FEATURES

source

1. 670
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE4152_B09_C18"
 /clone_lib="wheat CS whole plant cDNA library"
 /tissue_type="Roots, leaves, crown, stem and sheath"
 /dev_stage="Adult"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid
 pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; plant
 tissues from wheat cv. CS grown to full tillering stage in
 greenhouse were collected at University of California
 Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total

RNA was prepared from leaves (young leaf and third leaf),
 whole roots, crown, stem and sheath tissues, and then
 equal quantities of RNA were pooled from the these
 samples. PolyA was purified from the pooled RNA. A cDNA
 library was made, and the cDNA clones were in vivo excised
 to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.
 Akhunov, J. Dvorak) at the University of California,
 Davis. Colony plating, plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."

BASE COUNT 173 a 139 c 171 g 187 t

ORIGIN

Query Match 30.1%; Score 512; DB 14; Length 670;
 Best Local Similarity 85.7%; Pred. No. 1.7e-121;
 Matches 569; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 243 CTCAACACCCGCTGCGCAGCGGAGTCCCTCGGCTTCATGAGTCCAGCTGCTC 302
 Db 7 CAGAGGCTGGGCTGGCTGCTGAGCTCCCTCGGCTTCATGAGTCCAGCTGCTC 66
 QY 303 CTGCTCTCCATGAGCTCTCCCTCTGCTGCTCCTTACTTACTGAAATAGCATTT 362
 Db 67 CTGCTCTCCATGAGCTCTCCCTCTGCTGCTCCTTACTTACTGAAATAGCATTT 126
 QY 363 CTCTGAGGATGTTGCTGCTGAGTGTGTGATTAACAACCAAGATCAAGAAACA 422
 Db 127 CTCTGAGGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 186
 QY 423 AATGATGCTAATATAGCTTGAAGATAGATGTTGAACCAATGAGATTTTACCA 482
 Db 187 AATGATGCTAATATAGCTTGAAGATAGATGTTGAACCAATGAGATTTTACCT 246
 QY 483 GCTAGAGGACAGAGGAGTGTATTTGCTTAAGCTGATGTTATCTTAACACT 542
 Db 247 GCTAGAGGACAGAGGAGTGTATTTGCTTAAGCTGATGTTATCTTAACACT 306
 QY 543 GCTGTTGCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
 Db 307 GCTGTTGCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366
 QY 603 AAGATTTTGTGTGATCATGAATATGCTGAGGCACTTAAAGTTGAATATGTCAAA 662
 Db 367 AAGATTTTGTGTGATCATGAATATGCTGAGGCACTTAAAGTTGAATATGTCAAA 426
 QY 663 CATCTTCCTTTGTTGCTGAGGCACTGATTTATCTATCAACGCTGAGATTTGAT 722
 Db 427 CATCTTCCTTTGTTGCTGAGGCACTGATTTATCTATCAACGCTGAGATTTGAT 486
 QY 723 AGCAGACTAGGATCGCTGAAATATACAGATGTCACAACTTATGTTGTTACCTG 782
 Db 487 AGCAGACTAGGATCGCTGAAATATACAGATGTCACAACTTATGTTGTTACCTG 546
 QY 783 AATATGAAGATCTAATGAAAGTTGCTGAAGCAATGCTGCAAGAGAGTCTTGGGAA 842
 Db 547 AATATGAAGATCTAATGAAAGTTGCTGAAGCAATGCTGCAAGAGAGTCTTGGGAA 606
 QY 843 CATATTCGTAATCCCTTGAAGTACAGAGTGAAGATTCCTGTTGCAATTAATCACT 902
 Db 607 CATATTCGTAATCCCTTGAAGTACAGAGTGAAGATTCCTGTTGCAATTAATCACT 666
 QY 903 GTAT 906
 Db 667 GTTT 670

RESULT 7
 LOCUS BG464257 557 bp mRNA linear EST 20-MAR-2001
 DEFINITION EM1_71_E08_b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
 sequence.
 accession BG464257
 version BG464257.1 GI:13392578

KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 557)
AUTHORS Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@pratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 556
POLYA-No.

FEATURES
source
1..557
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr, Vector:
pBlueScript II from Lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from polyA RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 100 a 174 c 163 g 120 t
ORIGIN

Query Match 29.5%; Score 502.6; DB 12; Length 557;
Best Local Similarity 94.7%; Pred. No. 4,3e-119;
Matches 520; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 47 CCGGTCGACCGCGGATGCGCAAGACCCGTCGTCGCGGTGCGCGCGTCCGCGGAG 106
Db 9 CCGGTCGCGCGCGGATGCGCAAGACCCGTCGTCGCGGTGCGCGCGTCCGCGGAG 68
QY 107 CCGCGGCGCGGTTCACACCGGACCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 166
Db 69 CCGGCGCGCGGTTCACACCGGACCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 128
QY 167 CTCGGATCCACAGAGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
Db 129 CTCGGATCCACAGAGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 188
QY 227 TGGGAGACCCGCGCGCTTCACACCGGACCGGAGTCTCTCTCTCTCTCTCTCTCTCT 286
Db 189 TGGAGACCCGCGCGCTTCACACCGGACCGGAGTCTCTCTCTCTCTCTCTCTCTCT 248
QY 287 GTCCAAAGTCGTCT 346
Db 249 GTCCAAAGTCGTCT 308
QY 347 GATGSAATTGACATTTCTCTGAGGATGTTGGCTCGCAGTGGTGGTGAATGAACCA 406
Db 309 GATGSAATTGACATTTCTCTGAGGATGTTGGCTCGCAGTGGTGGTGAATGAACCA 368
QY 407 GAGATCAAGAAACAAATGATGATATGATGATGATGATGATGATGATGATGATG 466
Db 369 GAGATCAAGAAACAAATGATGATGATGATGATGATGATGATGATGATGATGATG 428
QY 467 AGTGCAGGTTTACAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 526
Db 429 AGTGCAGGTTTACAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 527 GGTATCTTAAACACTGCTGTGCTGCGAAAGTGCTTGAACCTGTTCTGAAGATCATGT 586

Db 489 GGTATCTTAAACACTGCTGTGCTGCGAAAGTGCTTGAACCTGTTCTGAAGATCATGT 548
QY 587 TCTTAAGT 595
Db 549 TCTTAAGT 557

RESULT 8
LOCUS BG268271/c 497 bp mRNA linear EST 20-FEB-2001
DEFINITION 1000172A07.x2 1000 - Unigene 1 from Maize Genome Project Zea mays
CDNA, mRNA sequence.
ACCESSION BG268271
VERSION BG268271.1 GI:12973018
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 497)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished (1999)
CONTACT: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1000172 row: A column: 07.

FEATURES
source
1..497
Location/Qualifiers
/organism="Zea mays"
/db_xref="dbEST:687063C09.y1"
/db_xref="taxon:4577"
/clone_lib="1000 - Unigene 1 from Maize Genome Project"
/note="This library represents the unique ESTs found in
the first round of EST sequencing at Stanford University
for the maize genome project. Sequences are present from
libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683,
, 687, 707, and 945. Contigs were assembled using TIGR's
CAP program and a representative EST from each contig was
selected."

BASE COUNT 122 a 147 c 111 g 117 t
ORIGIN

Query Match 28.9%; Score 492.2; DB 12; Length 497;
Best Local Similarity 99.4%; Pred. No. 2e-116;
Matches 494; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1133 AGCAATTGATGTCCTGTTGACAAATTCAGAGGCGGTGAGATGCTTTGGAAGATTAAC 1192
Db 497 AGCAATTGATGTCCTGTTGACAAATTCAGAGGCGGTGAGATGCTTTGGAAGATTAAC 438
QY 1193 AATTGAACATGGAATTCAGATGTCAGATGTTGGGACAGGCTGCGAGGACACGGA 1252
Db 437 AATTGAACATGGAATTCAGATGTCAGATGTTGGGACAGGCTGCGAGGACACGGA 378
QY 1253 GATGCTCTGAGGCGCTGCACTGCGCTTCTGCAATCTGCTGGAAAGAGGCGTGGCGCC 1312
Db 377 GATGCTCTGAGGCGCTGCACTGCGCTTCTGCAATCTGCTGGAAAGAGGCGTGGCGCC 318
QY 1313 TCTTGCAAGAAACATGTCAGATGTCGCAAGCAGCGGAGGAGGAGGCTCCATGAGGAG 1372
Db 317 TCTTGCAAGAAACATGTCAGATGTCGCAAGCAGCGGAGGAGGAGGCTCCATGAGGAG 258
QY 1373 AAAGGCTATGCGAGGCTGAGAGAAATGTTCAATGAGCAACCACTGGCTGAGGATGAC 1432

Db 257 AAGGCGTATGCGAGGGTGAAGAAATGTTCAATGAGACACCATGCTGAGAGGATCC 198

QY 1433 GCGCGTGTGAAGAGATGCTCTGAGAAATACAGAGACATCCAGGTCTTGAGCTTTGCC 1492

Db 197 GCGCGTGTGAAGAGATGCTCTGAGAAATACAGAGACATCCAGGTCTTGAGCTTTGCC 138

QY 1493 GTGCCCATCAGCTGCGCTACAGTGTGAATGATTTTACGGCTACGCTACCTGCTT 1552

Db 137 GTGCCCATCAGCTGCGCTACAGTGTGAATGATTTTACGGCTACGCTACCTGCTT 78

QY 1553 CAGCGCTAACTGATGATGACCTCTGTGTCTACTTTTTCACATTCATCTTTTACCT 1612

Db 77 CAGCGCTAACTGATGATGACCTCTGTGTCTACTTTTTCACATTCATCTTTTACCT 18

QY 1613 ATTAGGCGCATGCTCCAT 1629

Db 17 ATTAGGCGCATGCTCCAT 1

RESULT 9
B1959909 627 bp mRNA linear EST 22-OCT-2001
LOCUS HVSMEN0022F11f Hordeum vulgare rachis EST library HVCDNA0015
DEFINITION (normal) Hordeum vulgare cDNA clone HVSMEN0022F11f, mRNA sequence.
ACCESSION B1959909
VERSION B1959909.1 GI:16311164
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 627)
AUTHORS Wing, R., Close, T.J., Kleinof, A., Wise, R., Chin, A., Begum, D., Fritsch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R. and Main, D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 508
Seq primer: ATTAACTCTCACTAAAGG
High quality sequence stop: 612.
Location/Qualifiers
1. 627
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEN0022F11f"
/clone_lib="Hordeum vulgare rachis EST library HVCDNA0015 (normal)"
/rissue_type="Rachis"
/lab_host="TJCl21"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; site 2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinof lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Fritsch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Fritsch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and

contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinof A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/99pages/bgn/31/clover.html>).

BASE COUNT 162 a 135 c 185 g 143 t 2 others

Query Match 24.5%; Score 417; DB 13; Length 627;
Best Local Similarity 84.5%; Pred. No. 6,28-97;
Matches 479; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

QY 888 GCATATATTAACAGTGTATCAC GAGGAAAGGAGCAAGACTTATTTCTTCAAGCATTTTA 946

Db 18 GCTATATTAACAGTGTATCACGAGGAGGAGGAGCAAGACTTATTTCTTCAAGCATTTTA 77

QY 947 TCAGGCTTGACGCTCATCCACAGAGAACTTAAGTCTTACATATACATGCTGTAGT 1006

Db 78 TCAGGCTTGACGCTCATCCAAATCAGAGTGAAGTCCCAAGAGCATGCTGTAGT 137

QY 1007 TGTGGAAAGTATGTATGCTGTAGCCAAATTTGAATCTGACTTACTTGTGTGT 1066

Db 138 TGTGGAAAGTATGTATGCTGTAGCCAAAGGGGAGACACACTTACGTGAATTTGTGTG 197

QY 1067 GAAGAACAGATTCATGACAGCTGCTCATTTTGTGAACAAACATTTGGCAGTGGCCCTTA 1126

Db 198 GAGAAATGGAGATCCATGACCGTGTACCTTGTCAACAAACGTTGGCAGTGGCCTTA 257

QY 1127 CTGGGACGAATGATGTCTTGTGTGAATTTTCAGAGCCGCTGGAGAAATGCTTTGGAAG 1186

Db 258 TTGGCTGCAATTAATGTGCTTGTTCAGAACTTCAGCCCTGGAGAAATGCTTTGGAAG 317

QY 1187 GATTAACATTAAGCAATGACATTCATTAAGTGGCAGTATTTGGGACGCTGTCGAGGAG 1246

Db 318 GATTAACATTAAGCAATGACATTCATTAAGTGGCAGTATTTGGGACGCTGTCGAGGAG 377

QY 1247 CAGGAGATCTCTCTGAGAGGCTGCACTGAGCTTCTGATCTCTGAGGAGAGGAGGCT 1306

Db 378 CACGAAATGCTGTGATGTTCATACCGGCTTCTGACCTCTGAGGAGAGGAGGCT 437

QY 1307 GCGGCTCTTGGCAAGCAATCTGACTGCAAGCCAGCCAGAGGAGGCTTCAT 1366

Db 438 CAGGCTCTTGGCAAGCAATCTGACTGCAAGCCAGCCAGAGGAGGCTTCAT 497

QY 1367 GGGGAAAGGCTTATGCGAGGCTGAAGAAATGTTATGAGACCAATGCTGTAGAG 1426

Db 498 GGGAAACAAAGGCTTATGCGAGGCTGAAGAAATGTTATGAGACCAATGCTGTAGAG 557

QY 1427 GATCGGCGCGGTGTGAAGATGCTCT 1453

Db 558 CATCGCACGAGGTGTGAAGAGATGCT 584

RESULT 10
BG560599 439 bp mRNA linear EST 10-APR-2001
LOCUS RH122.59.D08.D1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
DEFINITION RH122.59.D08.D1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
ACCESSION BG560599
VERSION BG560599.1 GI:13589597
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 439)
AUTHORS Condomier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordemier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@uga.edu
Seq primer: JEN REV
High quality sequence stop: 429
POLA=No.

FEATURES
Location/Qualifiers
1. 439
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly-A RNA in the cloning vector Lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 100 a 107 c 123 g 109 t
ORIGIN

Query Match 23.6%; Score 401.6; DB 12; Length 439;
Best Local Similarity 95.6%; Pred. No. 5.4e-93;
Matches 413; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 206 GCGAGACCTTGGCCAGCGCGGTGGGAGCCCGCCCTTACACCGCGTGGCCAGCG 265
DB 8 GCGAGACCTTGGCCAGCGCGGTGGGAGCCCGCCCTTACACCGCGTGGCCAGCG 67
QY 266 GAGTCCCTCGGTTTCATGAGTCCAGCTGCTCTCTGCTCCCATGAGTCTTCCCT 325
DB 68 GAGTCCCTCGGTTTCATGAGTCCAGCTGCTCTCTGCTCCCATGAGTCTTCCCT 127
QY 326 CTGTGGTGGTCCACTTTTACTGATGAGATTAAGCATTTCTTGGAGCATTTGGCTTGA 385
DB 128 TTGGGGTGGCCACTTTTACTGATGAGATTAAGCATTTCTTGGAGCATTTGGCTTGA 187
QY 386 AGTGTGTGATTAACAACAGAGATCAACAAGAAACAAATGATGATATAGCTTGA 445
DB 188 AGTGTGTGATTAACAACAGAGATCAACAAGAAACAAATGATGATATAGCTTGA 247
QY 446 GCAAGAGATGTAACATGAGATGAGATTTTACAGCTAGAGACAGAGAGATTA 505
DB 248 GCAAGAGATGTAACATGAGATGAGATTTTACAGCTAGAGACAGAGAGATTA 307
QY 506 TATGCTTAAGCTGATCTGTTATCTTAACAACAGCTGCTGCTGCAAGTGGCTGA 565
DB 308 TATGCTTAAGCTGATCTGTTATCTTAACAACAGCTGCTGCTGCAAGTGGCTGA 367
QY 566 CCGTGTCTGAAGATGATGTTCCMAAGTCTTCCGAAATTTTGGTGGATCCATGA 625
DB 368 TCGTGTCTGAAGATGATGTTCCMAAGTCTTCCGAAATTTTGGTGGATCCATGA 427
QY 626 AATGCGTGGCA 637
DB 428 AATGCGAGGCA 439

RESULT 11
B0148371 685 bp mRNA linear EST 24-APR-2002
LOCUS B0148371
DEFINITION N067F07FL1063 Developing flower Medicago truncatula cDNA clone
B0148371
B0148371
B0148371.1 GI:20285430
ACCESSION
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula flower library
COMMENT Unpublished (2001)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert length: 685 Std Error: 0.00
Plate: 067 row: P column: 07
Seq primer: TCACACAGAAACGCTATGAC.

FEATURES
Location/Qualifiers
1. 685
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_lib="N067F07FL"
/clone_lib="Developing flower"
/tissue_type="Developing flowers"
/dev_stage="developmentally pooled. Contains a mixture of
very young, developing, fully-opened flowers and flowers
in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+
enriched, pooled samples of equivalent amounts of total
RNA from very young, developing, fully-opened flowers and
flowers transitioning into pods. The cDNA was
directionally ligated into the Uni-Zap XR vector
(Stratagene) and packaged using the GigaPack III Gold
packaging extracts. Phagemids containing cDNA inserts were
in vivo excised from the recombinant Uni-Zap XR vector
using Exsist helper phage and the E. coli strain
XL1-Blue MRF' (Stratagene). Excised plasmids were plated
using SDR cells."

BASE COUNT 202 a 115 c 162 g 203 t 3 others
ORIGIN

Query Match 23.2%; Score 395; DB 14; Length 685;
Best Local Similarity 74.8%; Pred. No. 3.2e-91;
Matches 505; Conservative 0; Mismatches 168; Indels 1; Gaps 1;

QY 510 GCTTAAGCTGATCTGTTATCTTAACAACAGCTGCTGCTGCAAGTGGCTT- GAACC 568
DB 12 GCTTAAGCTGATCTGTTATCTTAACAACAGCTGCTGCTGCAAGTGGCTGATGC 71
QY 569 TGTCTGAAGATCAAGTCTTCAAGTCTTCCGAAAGTTTGGTGGATCCATGAAT 628
DB 72 CTTCTTAAGAAAGATTACCTCGTTCTTCGAAAGTTTATGGATTCATGAAT 131
QY 72 CTTCTTAAGAAAGATTACCTCGTTCTTCGAAAGTTTATGGATTCATGAAT 131
DB 629 GCGTGGCATTTACTTAAGTGAATATGCAACAACATCTTCTTGTGCTGAGCAT 688
QY 629 GCGTGGCATTTACTTAAGTGAATATGCAACAACATCTTCTTGTGCTGAGCAT 688
DB 132 GCGAGGCAATTATTAAGAGCATATGCAAGCACTCCCTTTGTGAGGCTAT 191
QY 689 GATGATTTCTACACAGGCTGATGGAATAGCAGAGCTGAGGATGCTGAAT 748
DB 192 GATGATTTCTACACAGGCTGATGGAATAGCAGAGCTGAGGATGCTGAAT 251
QY 749 ACGATGCGCAAACTTATGTTGTTCACTGCGGAATAGTAAAGACTAATGAGTGC 808
DB 252 TAAATGCTGAAGACTATGTTGATCATCTTGAATAGCAAGAACTTATGGAAGTGC 311
QY 809 TGAAGCAATGTGCAAGAGAGTCTAGGGAACATTTGTGTAATCCCTTGAGTACG 868
DB 312 AGAAGATAGTGAAGAGAGGTTCTTGAGCACTTTGGAATCTTTGAGATAG 371
QY 869 GAGTGAAGATCTCTGTTTGCATATATTAACAGTATTCAGAGAAAGGCAAGACTT 928
DB 372 AATGATGATCTCTTTTGGCATCATTAATAGTGTTCAGGTGAAGGCAAGATCT 431

Tel: 516-344-3366
 Fax: 516-344-3407
 Email: burr@bnl.gov
 Seq primer: T3 Primer.
 Location/Qualifiers
 1..643
 /organism="Gossypium hirsutum"
 /cultivar="Acala Maxxa"
 /db_xref="taxon:3635"
 /clone_lib="Six-day Cotton fiber"
 /issue_type="mature fiber"
 /dev_stage="Six days post anthesis"
 /lab_host="X1L-blue"
 /note="Vector: pBluescript II KS+"

BASE COUNT 193 a 112 c 159 g 179 t
 ORIGIN

Query Match 21.2%; Score 360.4; DB 10; Length 643;
 Best Local Similarity 73.0%; Pred. No. 2.8e-82;
 Matches 463; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

FEATURES

source

430 TCACATATAGCTTGAGCATGATGTTGAACCATGAGTCAGGTTTACCACTAGAG 489
 2 TCATTACAGTTTAAACATAGATGTTGAACGAGGTTTCTCTGCCAAAG 61
 490 GACAGAGGCAAGTATGCTCTAAAGCGATCGTGTATCTTAAACAAGTCTGTTG 549
 62 GTCAGAGCACTATGATATGCTCTAAAGCGATGTTGATGTTGAACCTGCGTTG 121
 550 CTGGCAAGTGGCTTACCCCTGTTCTGAAGATCATGTTCTTAAAGCTCTTCCGAAGTTT 609
 122 CTGGAAGGCTGAGTGGCTGTTCTTAAAGCATGTTGCTGATGTTCTTCCAAAGTTT 181
 610 TGTGTGATGCCAATGATGCTGGGCAATTAAGTTGAATGTCAAACATCTTG 669
 182 TCTGTGATTCATGAATGAGGAGCATTTAAGTCAGAAATGTTAAGCACTCC 241
 670 CTTTGTGCTGAGCCATGATGTTCTCAACAAGCGCTGATTTGAATAGCAGCA 729
 242 CTTTGTGAGAGTCCATGATGTTCAACAAGCAAGCAATTAAGTGAAGCAAGCA 301
 730 CTAGCATGCTGGAATAACAGATCCCAAACTTATGTTTCACTGGGGAATGTA 789
 302 CTGGGAACGATTAAGATCAAAATGCTGAAACATAGTTTGACATTAGGAACAGCA 361
 790 AAGAACTATGGAATGCTGGAAGCAATGTCGAAAGATCCCAAGCAATATTC 849
 362 AGAGCTAATGGAATGCTGGAAGATGTCGAAAGATGTCGTAAGCATGTTT 421
 850 GTGAATCCCTTGAAGTACGAGTGAAGATCTCTGTTGCAATTAATTAACAGTATAC 909
 422 GTGAATCTCTGCTGGAATAAGATCTGTTGCTCTTAATTAATGTTTCCC 481
 910 GAGGAAGGGAACAAGTATTTCTTCAAGCAATTTTCAAGGCTTTGACGCTCAAC 969
 482 GAGGGAAGGTCAGACCTATTTCTCGTTCACTTTCATGAAGTTTGAATGATCAG 541
 970 ACAGAACTTAAAGTCTTAAGATACATCTGATGTTGGAGATGATGTTATGCTTC 1029
 542 AGAAGAACTTAAAGTCTCAACAATTCATGCGGATGATGAGAGTGCATGATGCTC 601
 1030 AGACCAATTTGAGACTCACTTACGTAAGTCTTGT 1063
 602 AACCAAGTTTGAACAGAGCTGCTGTAATCTTGT 635

RESULT 14
 AM257867 375 bp mRNA linear EST 30-MAR-2000
 LOCUS 687063C09.y1 687 - Early embryo from Delaware Zea mays CDNA, mRNA
 DEFINITION
 sequence.
 ACCESSION AM257867
 VERSION AM257867.1 GI:6626335

KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 375)

REFERENCE
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 687063 row: C column: 09.

FEATURES

source

1..375
 /organism="Zea mays"
 /cultivar="Illinois High Oil"
 /db_xref="taxon:4577"
 /clone_lib="687 - Early embryo from Delaware"
 /issue_type="embryo"
 /dev_stage="14, 21, 28, and 35 days after pollination"
 /lab_host="E. coli SOLR"
 /note="Organ: embryo; Vector: pBluescript SK; Site.1: XhoI
 ; Site.2: EcoRI; Library was prepared by Scargene using
 the Uni-ZAP XR system (Stratagene BN37328-12). Clones
 were picked by a Q-bot after blue/white selection
 (ampicillin resistance - use 100 micrograms/microliter).
 Developed from a pool of equal amounts of RNA from
 developing embryos sampled at 14, 21, 28 and 35 days after
 pollination of the Illinois High Oil Maize Strain Cycle
 90. This closed strain has been selected for high oil
 concentration for 90 generations and originates from the
 1890s era open pollinated variety Burr's White"
 80 a 104 c 111 g 77 t 3 others

ORIGIN

Query Match 20.1%; Score 342.4; DB 10; Length 375;
 Best Local Similarity 96.0%; Pred. No. 1.1e-77;
 Matches 360; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

155 GGCCTGCGACCTCCGCAATCCACAGAGGTTCTCTCCGCGGTCCCTCGAGACCC 214
 1 GGCCTGCGCAACCTCCGCAACCAAGAGGGTCTCTCCGCGGAGCCCTCGAGACCC 60
 215 TT-GCGAGCGCGTGGGAGACCCCGCGCCCTCAACACCGCGTCCGCAAGCGAGTCCC 273
 61 TTGGGAGCGCGTGGGAGACCCCGCGCCCTCAACACCGGNGTCCGCAAGCGAGTCCC 120
 274 TGGGTTTANGAGTCCAGCTGCTGCTCTCTCCCAAGACTCTCTCTCTGGTG 333
 121 TGGGTTTANGAGTCCAGCTGCTGCTCTCTCTCCCAAGACTCTCTCTCTGGTG 180
 334 GTCCACTTTTATGATGAAATGATTTCTTCAAGCATGTTGGCTCGCAAGTGTGT 393
 181 GTCCACTTTTATGATGAAATGATTTCTTCAAGCATGTTGGCTCGCAAGTGTGT 240
 394 GATTAACCAACCAAGATATCAAGAAACAAATGATGTCACTATAGTTGGAGCATAGA 453
 241 GATTAACCAACCAAGATATCAAGAAACAAATGATGTCACTATAGTTGGAGCATAGA 300
 454 TGTGAACCAATGAGTGCAGTTTACCACTAGAGGAGAGGAGGAGTATATGCTC 513
 301 TGTGAACCAATGAGTGCAGTTTACCACTAGAGGAGAGGAGGAGTATATGCTC 360
 514 TAAAGCTGATCTGG 528
 361 TAAAGCTGATCTGG 375

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132526.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139497.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140355.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142155.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143544.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144088.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147934.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148349.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151530.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160999.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	65.2%;	Score 1563.5;	DB 21;	Length 457;
Best Local Similarity	69.0%;	Pred. No. 1.8e-152;		
Matches 309;	Conservative 63;	Mismatches 65;	Indels 11;	Gaps 2

QY	29	LVAVAAVSASTGFLTGLCARIDPC-----DGRGDPAAINTAVASGSPLEFNRSLKYV	80
Db	8	LMWLITLSTVSTVCVILVRSSFECSSISQPFYERKNGESA---AKQSNPMDPMKSTLV	64
QY	81	LVSHELSISGGBLLIMELAFRLRHVGSOVVTINQSGETNDVTYSLEHRLMHNQVOYL	140
Db	65	LVSHELSISXHPLELMELAFLLRGVADVVWITNQPLREDEVEVYSLHEKMLDRGVOYIS	124
QY	141	ARGGEANDTALKADIVILTNAAVAGKMLPVTYKDHVKYV.PKILWVHEHKGHYFVEVYK	200
Db	125	AKGGKAVDTSLKADIVILTNAAVAGKMLDAVKEVNAKVYV.PKILWVHEHKGHYFNAADV	184
QY	201	HLPEVAGAMIDSHSTTAAYVNSRSTDRKLTQMPQTYVYVHLSGSKELMEVADNVAARVLR	260
Db	185	HLPEVAGAMIDSHATAGVWKRTQARLGIKMPKYVYVHLSGSKELMEVADNVAARVLR	244
QY	261	HIRESLCYNSBULLPAILINSVSRGKGODLFLQAFQALQLOHKLKVPRIIAVYVNSDV	320
Db	245	HVRESLGVYNEDELLFGIINSVSRGKGODLFLRAFHESLERLKEKKLOVPMIAVYVNSDM	304
QY	321	NAQCKPEFQOLARDPVVUKNTIDHRVHPFNKNTLVAVYLAIDVLYONSGRGCEFGRTTEA	380
Db	305	SKQCKPEBELRNPAREKKLENFVHPFNKTLVAVYLAIDVLYONSGARGCEFGRTTEA	364
QY	381	MAFKLPIVLTAAAGTTEIVLDGSTGLHPAGKEGVAPLAKNIVLASHAQRVNSGKEY	440
Db	365	MAFKLPIVLTAAAGTMEIVNGTGLIHSAAKEGVPLAKNIVYLAATQVETLRLMGKGY	424
QY	441	GRVKEKMEHHAPELIAVYKDVLRKSQ	468
Db	425	ERVKEKMELEHSHSRILASVLEKVLQAK	452

XX	RESULT 2
XX	AAAG28810
XX	ID AAAG28810 standard; Protein; 463 AA.
XX	AAAG28810;
XX	17-OCT-2000 (first entry)
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 34168.
XX	Protein identification; signal transduction pathway; metabolic pathway;
XX	hybridisation assay; genetic mapping; gene expression control; promoter
KM	

KM	termination sequence.	
XX	Arabidopsis thaliana.	
OS	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
PF	25-FEB-2000; 2000EP-0301439.	
XX		
XX	25-FEB-1999;	99US-0121625.
PR	05-MAR-1999;	99US-0123380.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	08-APR-1999;	99US-0128234.
PR	16-APR-1999;	99US-0128714.
PR	19-APR-1999;	99US-0129845.
PR	21-APR-1999;	99US-0130077.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132663.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	16-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134991.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135639.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137282.
PR	03-JUN-1999;	99US-0137558.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138034.
PR	10-JUN-1999;	99US-0138074.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139432.
PR	16-JUN-1999;	99US-0139433.
PR	17-JUN-1999;	99US-0139464.
PR	18-JUN-1999;	99US-0139465.
PR	18-JUN-1999;	99US-0139465.
PR	18-JUN-1999;	99US-0139477.
PR	18-JUN-1999;	99US-0139478.
PR	18-JUN-1999;	99US-0139479.
PR	18-JUN-1999;	99US-0139480.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139879.
PR	22-JUN-1999;	99US-0140381.
PR	23-JUN-1999;	99US-0140383.
PR	24-JUN-1999;	99US-0140635.

[illegible]

PR	10-SEP-1999	99US-0153070
PR	11-SEP-1999	99US-0153078
PR	15-SEP-1999	99US-0154018
PR	16-SEP-1999	99US-0154039
PR	20-SEP-1999	99US-0154779
PR	22-SEP-1999	99US-0155139
PR	23-SEP-1999	99US-0155446
PR	24-SEP-1999	99US-0155659
PR	28-SEP-1999	99US-0156488
PR	29-SEP-1999	99US-0156596
PR	04-OCT-1999	99US-0157117
PR	05-OCT-1999	99US-0157573
PR	06-OCT-1999	99US-0157865
PR	07-OCT-1999	99US-0158029
PR	08-OCT-1999	99US-0158222
PR	12-OCT-1999	99US-0158322
PR	13-OCT-1999	99US-0159233
PR	13-OCT-1999	99US-0159294
PR	13-OCT-1999	99US-0159295
PR	14-OCT-1999	99US-0159330
PR	14-OCT-1999	99US-0159330
PR	14-OCT-1999	99US-0159637
PR	14-OCT-1999	99US-0159638
PR	18-OCT-1999	99US-0159854
PR	21-OCT-1999	99US-0160711
PR	21-OCT-1999	99US-0160767
PR	21-OCT-1999	99US-0160770
PR	21-OCT-1999	99US-0160814
PR	21-OCT-1999	99US-0160815
PR	22-OCT-1999	99US-0160981
PR	22-OCT-1999	99US-0160981
PR	22-OCT-1999	99US-0161049
PR	25-OCT-1999	99US-0161404
PR	25-OCT-1999	99US-0161406
PR	26-OCT-1999	99US-0161359
PR	26-OCT-1999	99US-0161360
PR	28-OCT-1999	99US-0161920
PR	28-OCT-1999	99US-0161920
PR	28-OCT-1999	99US-0162193
PR	29-OCT-1999	99US-0162142

Query Match	65.2%;	Score 1563.5;	DB 21;	Length 463;
Best Local Similarity	69.0%;	Pred. No. 1.8e-152;		
Matches · 309;	Conservative 63;	Mismatches 65;	Indels 11;	Gaps 2

```
QY      LVVAASVSTINGELGALRDC-----GQRDPAINTAVAGSGLTGMSKLYL 80
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      LMVLFEISVTVCULVRSSPFCSCSISSQPVEENKESGA---AKFOSNLPFMKSPLYL 70
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      LVSHETLSGSGPLLMELAFLRHVSQGVWIMINQSOGTNDJYXSLERHMLNHGOVLP 140
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      LVSHETLSGXPELLMELAPLRHGVALVVWITNQPRLEDESVYSLEHHMKLDREGVOVIS 130
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      ARGEADVIALKADVLILNTAIVAGKMLDPVLKHVKVYLPKLIWMIHBMKGHYFCEVYX 200
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      AKGGKAVDTSELKADILVLTANTAAVGKMLDAVLKENNVKVLPKLIWMIEHRMGHYFNADLVK 190
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      HLPFVAGAMTDSHTTAAYMNSTRSDRLTIQMPQCTYYHLGNSKELMEVMDVARVURE 260
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      HLPFVAGAMTDSHTTAAYMNKNRTQARLTIKMPTYYVHLGNSKELMEVMDVARVURE 250
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      HIRSESVASBEDLFAINSVSRGQQLFLFOAOLIQHEKLKYPRTHIAVVVGSDV 320
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      HVRSISGVANNEDLFGLINSVSRGQQLFLRAFHSLERIKEKKQLVPHTMAVVVGSDM 310
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      NAQCKEFTQARPVAVKQITHDRHNFPVKTAAVAPYLAIDVLVNQSGRGCEFGMITTEA 380
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      SKQCKEFTELRNFRREKKLENFHPVNNVTTLVAPYLAIDL VNQSGARECGFITTEA 370
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


'Tue Jul 8 09:50:07 2003

us-09-938-294-45.rag

Page 5

Qy	361	MAFLPLVLGTAAAGCTTEIVLDGSTGLHPACKESGVAPLAKNIIVRLASHAEQVRSNGEKY	440
Db	371	MAFLPLVLGTAAAGCTTETIVNGTTGLHSAKKEGVIPLAKNIIVKLAQVELRLMKNGY	430
Qy	441	GRVEMFMEEHMAERIALVAKDVLRSQ	468
Db	431	ERVCEMFLHHMSRIASVLYKEVLOHAK	458
RESULT 3			
ID	AAG28812		
AC	AAG28812 standard; Protein; 449 AA.		
XX	AAAG28812;		
DT	17-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 34170.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KM	termination sequence.		
XX	Arabidopsis thaliana.		
OS	Arabidopsis thaliana.		
PN	EP1033405-A2.		
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130891.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132484.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	06-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135114.		
PR	21-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136392.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		
PR	04-JUN-1999; 99US-0137562.		
PR	07-JUN-1999; 99US-0137724.		
PR	08-JUN-1999; 99US-0138094.		
PR	10-JUN-1999; 99US-0138540.		
PR	10-JUN-1999; 99US-0138847.		

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0150560.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158359.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.1%; Score 1559.5; DB 21; Length 449;
Best Local Similarity 68.9%; Pred. No. 4.5e-152;
Matches 308; Conservative 63; Mismatches 65; Indels 11; Gaps 2;

QY 30 LVVVAASASTAGFLRKAALDDPC-----DGRGDPAINTVAVSGSPGLGPMMSKIVYL 81
DB 1 MVLPLPSTVCTVILVRSFSCSISQFVEKNQSSA---AKQSNPLDFMKSKIVYL 57

QY 62 VSHRLISGGPLLIMELAFILRHVGSQVWINTNRSOETDVYTSLEHRLNMGVOYLA 141
DB 58 VSHRLISXXFLLIMELAFILRHVGSQVWINTNOKPLEDEBIVYTSLEHRLMDRGQYISA 117
QY 142 RGEAVDIAIKADVILNTAVAGKMDPVLKDHVPKVLPIKLWIHEMRGHYFKVEYVKH 201
DB 118 KQQRADVDTSLKADLLVLTAVAGKMDAVLKENVVKKVLPKLWIIHEMRGHYFNADLVKH 177
QY 202 LPFVAGMIDSHHTAEVNSRSTSDRLKIOMPTQYVHLGNSKEAMEVAEDNVARVIREH 261
DB 178 LPFVAGMIDSHHTAGYKRNRTQARLGICMPTQYVHLGNSKEAMEVADSAAKVVIREH 237
QY 262 IRESLVGRSEDLIPATINSVSRGGQDLFLQAFYALQLOIHEKLVPIIAVAVVGSVN 321
DB 238 VRESLGVNRNDLLFGIINSVSRGGQDLFLRAFHESLERIKERKLDQVPMHVVVGSVMS 297
QY 322 AQTKEETQURDPVVKNTIHDVHFVNKTLAVAPYLAIDVLYVNSQGRGECFGRITTEAM 361
DB 298 KQTFETELRNPFVRKQLENFVHFVNKTLVAPYLAIDVLYVNSQGRGECFGRITTEAM 357
QY 382 AFKLPLVLTGTAAGTTEIYLDGSTGLHPAGKEGVAPLAKNIVRLASHAQRVSMGEGYX 441
DB 358 AFKLPLVLTGTAAGTTEIYVNGTGLLHSAKEGVIPLAKNIVRLATQVELRLRMKGNGYE 417
QY 442 RYKEMFMEHMAERIAAVLKDVLRKSQ 468
DB 418 RYKEMFLEHMSHRIASVLRKEVLQHA 444

RESULT 4
ABB92331 standard; Protein; 697 AA.
ID ABB92331
XX ABB92331;
AC ABB92331;
XX 31-MAY-2002 (first entry)
DT 31-MAY-2002
XX Herbicidally active polypeptide SEQ ID NO 1542.
DE Herbicidally active polypeptide SEQ ID NO 1542.
XX Herbicidal; plant; agriculture; herbicide.
KW Herbicidal; plant; agriculture; herbicide.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN MO200210210-A2.
XX 07-FEB-2002.
PD 07-FEB-2002.
XX 28-AUG-2001; 2001MO-EP09892.
PF 28-AUG-2001; 2001MO-EP09892.
XX 28-AUG-2001; 2001MO-EP09892.
PR (FARB) BAYER AG.
XX Tietjen K, Weidner M;
PI WPT; 2002-269010/31.
DR WPT; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX Claim 5; SEQ ID NO 1542; 261pp + Sequence Listing; English.
XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

SQ Sequence 697 AA;

Query Match	20.3%;	Score 487;	DB 23;	Length 697;
Best Local Similarity	29.1%;	Pred. No. 7.5e-41;		
Matches 147;	Conservative 77;	Mismatches 159;	Indels 122;	Gaps 13

QY	5	COORGPALMTAAVASGPIGPMR---	SKVLIVSHETLSGCPILMETAFILHVSQ	1087
Db	223	CRKSD-----	FKRLVMSRRVLLPFELMTGAPISMMETASLLSGAT	267
QY	109	VWITNOSGCTNDVTYSLEHMLNHGQVLPARGOAVDIALADLYILNTANAGKLD	1668	
Db	268	VAAVULSRG-----	GLQELTRRRKQVEDGSLSTKTMKADIVLIGSVQASMTD	320
QY	169	PLVKDHPKVLPKILMWIHEMGHYFK-----	VEYKHLPPVAGAMIDSHTAAYWNS-R	2222
Db	321	QVM-DHNPAGSGQIAMVWVNNREYFEDAKPVLDRVKLLIFL-----	STVOSKQMLTWC	3737
QY	223	TSRLKIQMPTQYVHLHNSKELMEVAEDVA-----	RVLRHIRESLGV	2686
Db	374	EDHVKYLR-SQPVVPLSVNDLAFVAVGSSSLNFTPLTQETPMKCKQKRLRSVATTEGL	4322	
QY	269	RSBDLLFALINSVRGQDPLFLCAFYQAL-----	IQHETLK	307
Db	433	TDMQMLVMSLSSINPGQULLLESVALTEREQTOEOVARKNOSKITIKNLNGIRKETS	4922	
QY	308	VPRIHAV-----	VVSGDVN	321
Db	493	LSARHRLGSSRKMKITSPAVDNHPSVLSATGRKLLLSGNVTOKODKILLLSGVGSSKN	5522	
QY	322	AQCFEFTOLRPFVVK-N-TIDHRVHFNKTLVA-PYLAIDVLMONSOGRECFGITTEA	380	
Db	553	-KVAAYKEMLSFLSNQNLNSVLTWPATTEVSLSADYVYVNSQGVETPGRVTIEA	6111	
QY	381	MAKRLPVLTAGGTETIVLDGSLGILAPRAGKEGVA-PLAKNIVYLRASHABOPVSGEYQ	4404	
Db	612	MALGVLGVLTGDSGTKEIVHNYTGLILPVRAGNKVLYQNLILFLRNPNSTRLOLQSGOR	6711	
QY	441	GRVXEMPEHMAERIAVALIKDLR	465	
Db	672	EIVKRYMKOHMYKRFVDVLVKCMR	666	

	RESULT 5
AB91330	
ID	ABB91330 standard; Protein; 670 AA.
XX	
AC	ABB91330;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Herbicideally active polypeptide SEQ ID NO 541
XX	
KW	Herbicidal; plant; agriculture; herbicide.
XX	
OS	Arabidopsis thaliana.
XX	
PN	MO200210210-A2.
XX	
PD	07-FEB-2002.
XX	
Pf	28-AUG-2001; 2001WO-EP09892.
XX	
PR	28-AUG-2001; 2001WO-EP09892.
XX	
PA	(PARE) BAYER AG.
XX	
PI	Tietjen K, Weidler M,
XX	
DR	WFI ; 2002-269010/31.
XX	

PT -identifying plant target proteins for herbicidally active compounds,
PT -comparing aligning and comparing nucleic acid or amino acid sequences
PT -from plant with nucleic acid or amino acid sequences from non-plant
PT -organisms -

PS Claim 5; SEQ ID NO 541; 261pp + Sequence Listing; English

CC The invention relates to identifying target proteins
CC (AB990790-AB990416) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

5Q Sequence 670 AA;

Query Match	20.1%;	Score 482.5;	DB 23;	Length 670;
Best Local Similarity	29.2%;	Pred. No. 2.1e-40;		
Matches 148;	Conservative 77;	Mismatches 173;	Indels 109;	Gaps 13;

```

0Y 38 STAGFELLR--GALBD-----PCGRGRDPAALNTAVAGSFLGPMK---SKVL 80
Db 193 ATYGKLLRPFSGLEKVLKEMSPHRRSCTGCRKSD-----FKLVMSRFFV 237
0Y 81 LVSHSLSSGGGLLMLBLAFLLRHVGSQVWVITNQSDETDVYLSLHRMLNHQVLP 140
Db 238 LLFHELSMTGAPILSMELASSELLSCGATVSAVVLISBRG-----GLMOELSRRIKVE 290
0Y 141 ARGEAVDILAKADVLINLVAAGKMLDPVLQDHVKVLKILMIWHMRGNYCF----- 195
Db 291 DKGSLSTKAMKADILLIISAVCTSMIQVNNHH--PAGSQJAMMIENRREYFPAKAV 349
0Y 196 VEYKHLPEVACAMIDSHTTAEVYNS--RTSDRLKIQMPQTVYHLGNSKELMEVA----- 249
Db 350 LDRVKMLIFL-----SEGSQRMLTWCEBEHIIKL-SQPVIVELSVNDELAFAVGPIS 402
0Y 250 -----EDNVARVLRRHIFRESIGVRSDDLFAIINSRGRGDLFLQAFQALQ- 299
Db 403 LNTPLSPKMKVKKQILRESVTELGITSDMLVMSLISNPTKGILLSLIALISE 462
0Y 300 -----LIQHEKILKVPRIHAV-----VVGS 318
Db 463 RGEOSQRNHKGIIRREKAVSLSSKHLRGSSSHQKSVSLTLNDGLRREKQELKVLGVS 522
0Y 319 DVNAQIKFETQLRDPVYKANTIHDRVHPFNKTLVAAYLYLALDVLVQNSQREBEGRTI 378
Db 523 KSNVGVYKEMSLFSLNSGNLSKSVMMIPATTRAAVLYSAAVYVYVNSQCVETGRATL 562
0Y 379 EAMAFLKPYLGTAAGGTTEIVLDGSTGLHFPAGKEGAVALAKIVRLASHASROVSMGEK 438
Db 583 EAMA YGLAVVGDAGCKREKVQNMHTGHLISMGSNKRKLAINLILYLRNRPDRLGSE 642
0Y 439 GYGVYKEMPEMEHNAERIALVLDKVL 465
Db 643 GRKXVERMTKQAMKRRFVILVKR 666

```

RESULT 6	
ABB91331	
ID	ABB91331 strand; Protein; 670 AA.
XX	
AC	ABB91331,
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Herbicideally active polypeptide SEQ ID NO 542
XX	
KW	Herbicidal; plant; agriculture; herbicide.
XX	
OS	Arabidopsis thaliana.

	PN	MO200210210-A2.
	XX	
	XX	PD 07-FEB-2002.
	XX	
	PF	28-AUG-2001; 2001WO-EP09892.
	XX	
	PR	28-AUG-2001; 2001WO-EP09892.
	XX	
	PP	(FAR) BAYER AG.
	PA	
	XX	
	XX	Tierjen K. Weidler M;
	PI	
	PT	WPI; 2002-269010/31.
	DR	
	PT	Identifying plant target proteins for herbicidally active compounds,
	PT	comprising aligning and comparing nucleic acid or amino acid sequences
	PT	from plant with nucleic acid or amino acid sequences from non-plant
	PT	organisms -
	XX	
	PS	Claim 5; SEQ ID NO 542; 261pp + Sequence Listing; English.
	CC	
	CC	The invention relates to identifying target proteins
	CC	(AB990790-AB994016) for herbicidally active compound, comprising
	CC	aligning and comparing nucleic acid or amino acid sequences from plant
	CC	with nucleic acid or amino acid sequences from non-plant organisms using
	CC	suitable search parameters, where plant sequences having an E-value
	CC	greater by a factor of 3 than the E-value of most similar non-plant
	CC	sequences are selected. The polypeptides or nucleic acids encoding them
	CC	are useful for identifying modulators. The identified modulators are
	CC	useful as herbicides.
SQ	Sequence	670 AA;
Query Match	20.1%; Score 482.5; DB 23; Length 670;	
Best Local Similarity	29.2%; Pred. No. 2.le-40;	
Matches 148; Conservative % 77; Mismatches 173; Indels 109; Gaps 13		
OY	38 STAGFLLR--GALRD-----PCORGDPAALNTAVASGSPLGFM--SKVLV 80	
DB	193 ATTGKLGCPGSLDEKYLEMSPHRSGTCDKRSD-----FKLWSSRFV 237	
OY	81 LVSHELTSSGGPILMLLEAFLLRHVSQVVWITNORSDETNDVTYSLEHRLMHNGOVL 140	
DB	238 LTFHELSMTGAPISMMELABELSCGATVASVAIVLSRG-----GLMGELSRRRIKIVE 290	
OY	141 ARGGEVDIAAKADLVILNTAAVAGKMDLPYLKHGVKXULPKILWIHNHGRGYFR-- 195	
DB	291 DKGELSFKTAMKADIITIASNACTSNIDQYNHH--PAGSQIAMIENREKREYPDRAKV 349	
OY	196 VEYVKHLPFVAGAMIDSHTTAETYNWS--RTSDRLKIQMPOTVYVHGNSKEIMEVA---- 249	
DB	350 LDVRKMILPL-----SESQSQRWLTCEEHEHIKL-RSQPYIVPLSVNDELFAVAGIPSS 402	
OY	250 -----EDNVARVLREHIRELYGVSDDLPIITINSVRKGODFIQAFVOLQ-- 299	
DB	403 LANPTLPSPEKKRVKQOLIRRSVRTBLGITDSVMLSNSTINPTKQDLLLESIALSE 462	
OY	300 -----LIQHEKLVKPRIHAV-----VVGSI 318	
DB	463 RGOESQRNHKGIIRKEKVISLSSKRRLRGSSROKMSVSLTDNGLRREKOKELVLLGSVGS 522	
OY	319 DVNAQCFKFPTQLRPFVKNKITIDRHVPNKTLVLAAYUAILDIVONSGRGECFGRTI 378	
DB	523 KSKHKGVYKEMLSFLSNSGNLSKSWMTPATRYVALSADVYVYTNSSQVEFTFGKVTI 582	
OY	379 ENAAFCLPVLGTGAAGCTTEIVLDGSTGLIHDPAGEGVAPLAKNIIVRIASHAQRVSNGEK 438	
DB	583 ENAAVGLAVAGTDDAGGTRKENVQHANNVTGLLSMGRSGNKEIAHNLVYLLRNPPDERLRIGSE 642	
OY	439 GYGRVKEVMEHEHHAERITAAVLKDVL 465	
DB	643 GRKQVEKQYMKOHMYKKFPYDVLYVKMR 669	

```

1  RESULT 7
2  ABB57933
3  ID ABB57933 standard; Protein; 424 AA.
4  AC
5  XX ABB57933;
6  DT 26-MAR-2002 (first entry)
7  DE Drosophila melanogaster polypeptide SEQ ID NO 591.
8  DE Drosophila melanogaster polypeptide SEQ ID NO 591.
9  KW pharmaceutical.
10 KW
11 KW Drosophila melanogaster.
12 OS
13 PN MO200171042-A2.
14 PD
15 XX 27-SEP-2001.
16 PE
17 XX 23-MAR-2001; 2001MO-US09221.
18 XX
19 XX 23-MAR-2000; 2000US-191637P.
20 PR 11-JUL-2000; 2000US-0614150.
21 PA
22 XX (PEKE ) PE CORP NY.
23 PI
24 XX Venter JC, Adams M, Li PMD, Myers EW;
25 DR N-PSDB; ABL02036.
26 DR
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
9
```

QY 276 AIIINSVSRGKODLFLQAFYQALQIOHEKLVKPIHAAVVG-----SDVNA 322
 DB 226 LDIRRYERKKGNALHLSRLIGDMLPADYFK--RCRLIINGAGYDRCMENVEHAELEH 283
 QY 323 QTKETQLRDFVY-----KNTIHDRVHFVNKTLAVAPYLAIDVLYONSQSGECFGRIT 377
 DB 284 LTF-ELKLDHVVLLRSPDEKCRLLFAHCLLYTP-----ENEHFGLVP 328
 QY 378 IEAAAFPLPYLGTAAAGTTEIVLDGSTGLHPAGKEGVAPLAKNIYVLAHAEQVSMGE 437
 DB 323 LEEGYCKRFVVALNSGSPETVYVSTSTGFLCEKTEKS---FGAMQLPRDQLLVKGD 385
 QY 438 KGYGVKEMFEMEHMAERIALAVLQVLRKSQEH 471
 DB 386 QGHKRVQOKFSQAFADRLNGIIRDLDVIRRESS 419

RESULT 8
 AAY68959
 ID AAY68959 standard; Protein; 385 AA.
 AC AAY68959;
 DT 30-MAY-2000 (first entry)
 XX Cps2g protein which has glycosyltransferase activity.
 DE Cps2g protein which has glycosyltransferase activity.
 XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
 KW capsular component; antigen; regulation; chain length determination;
 KW complement-mediated opsonophagocytosis; serotype-specific detection;
 KW antigen; vaccine; Streptococcal disease; ORF 22; ORF 21; ORF 22;
 KW Cps2A; Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J;
 KW Cps2K; Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.
 XX Streptococcus suis.
 OS Streptococcus suis.
 XX
 FH Key location/Qualifiers
 FT Misc-difference 266 /note="unspecified amino acid encoded by AGA"
 FT
 XX W0200005378-A2.
 PN 03-FEB-2000.
 PD 19-JUL-1999; 99MO-NL00460.
 PF 22-JUL-1998; 98EP-0202465.
 PR 22-JUL-1998; 98EP-0202467.
 XX
 PA (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
 XX
 PI Smith HE;
 XX
 DR WPI; 2000-195104/17.
 DR N-PSDB; AA260929.
 PT New nucleic acid containing the capsular gene cluster of Streptococcus
 PT suis, used for serotype-specific detection and to generate antigens or
 PT mutants for vaccination
 CC
 PS Disclosure; Fig 3; 144p; English.
 CC The proteins AAY68950-69 are encoded by the capsular gene cluster of
 CC Streptococcus suis serotype 2. The genes in this cluster are involved
 CC in polysaccharide biosynthesis of capsular components and antigens. The
 CC proteins are involved in regulation (CpsA), chain length determination
 CC (CpsB, CpsC), export (CpsD), and bioynthesis (CpsE, CpsF, CpsG, CpsH,
 CC CpsJ, CpsK). The capsule confers bacterium resistance to
 CC complement-mediated opsonophagocytosis. The gene cluster is used as a
 CC source of probes and primers for serotype-specific detection of S. suis
 CC and is also useful for recombinant production of the proteins. The
 CC proteins are then useful for producing antigens that can be used in
 CC vaccines, for controlling or eradicating a Streptococcal disease, in

CC humans or animals, e.g. against S. suis in pigs.
 XX
 SQ Sequence 385 AA;
 Query Match 6.7%; Score 161.5; DB 21; Length 385;
 Best Local Similarity 24.4%; Pred. No. 1.3e-07;
 Matches 90; Conservative 61; Mismatches 145; Indels 73; Gaps 20;
 QY 102 LRHVSGQVVMITNQ--RSQETN-----DVTYSLERMLNHGVVLPARGQEAVDIALKAD 154
 DB 48 LREVGQVVEIVNYPILRRKYFNPKGIFDYFISYHH---YSKOI-----AQVAIE--NKVD 97
 QY 155 LVLIINT--AVAGKMIDPYLKDHPKYLKILMWTHERGHYFVEYVK--HPFVAGAMID 211
 DB 98 ITHNNTTAVLEGIYKRKLK-----LP-LIMVHVEI--IYKPKISDSINFLMGCFAD 147
 QY 212 SHTTAEVMSRTSDLK-----IOMPQTYVVLHNSKMEVADNVARRYREHIRESLG 267
 DB 148 KIVTV---SQAVAHNIKQSPHIDQISVIYGVNKFVYQSD---ARSV-----RERFD 196
 QY 268 VRSEDLFALINSVSRGKODLFLQAFYQALQIOHEKLVKPIHAAVVGSPVNAQ----- 323
 DB 197 IDEEALVIGMGRVANKGQDFLEAVPILF-----QNPRAINFAGSAFEGEMRV 249
 QY 324 TKFETQLRDFVYKNTIHDRVHFVNKTLAVAPYLAIDVLYONSQSGECFGRITTEMAP 383
 DB 250 VELEKTIQSLKVSQVXKMDYVANTT---ELNMVFDFVLPST--NPDELPTVYLKAMAC 304
 QY 384 KLPLYGTAAAGTTEIVLDGSTGLHPAGKEGVAP-----LAKNIYVLAHAEQVSMGEK 438
 DB 305 GKPFVGRIGRGCEVKEGVNGFL-----VTNPSPLNLSKYLQDSNINILKRIKIGN 357
 QY 439 GYGVKEMF 447
 DB 358 SIERKKEHF 366

RESULT 9
 AAG43119
 ID AAG43119 standard; Protein; 405 AA.
 AC AAG43119;
 DT 18-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 53657.
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 53657.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 PD 06-SEP-2000.
 PF 25-FEB-2000; 2000EP-0301439.
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0132486.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134761.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143552.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148371.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

Oy	76	SKVLVYSH-ELISGGEDLLMEIAFLIRHVSQV-VWITNQRQEDNIVYLSLHEMLN	133
Db	7	SKNNMLITHPDLGIGGERLIVDAAVELASHGHKVAHIFSHDXSRCFEETLSGIFQVYV	66
Oy	134	HGVQVLPAR--GQEAVDIALKADLVILNTAVAGKMLDEVILKDHPRKVLK-----KI	182
Db	67	YG-SFLPRHI FYRLHAAVCAYLRCLFVALCVTLGMSFSDVVLADVSVPVPLKLRSSKY	125
Oy	183	LMWH-----EMRGHYEK-VEYVKLPFVAGAMI--DSHTTAEMNSRTSDRLK	228
Db	126	VFYCHPFDLLAKHTTLTKRMTRKPIDEIEQTTGMAADMLVNSNFTASTF-AATFKRLN	184
Oy	229	IQMFQTYVYHIGNSKEIMEVAEDNVARRYLREHIRESLGVSDELLFATINSVSKGQD	288
Db	185	AQGRPAVLYPA-----VNIDQIFEPHYTKINELSLNFEERKNID	225
Oy	289	LFIDAPQALDQIQHEKIKATPRIHAAVVG-----SVNMQTFEYQDADRVKNTIHDRV	343
Db	226	LAVSAF-----ALCKHQNLSDVTLTVAGKGYDEKKNENYLEELRSLAKEGVSRV	281
Oy	344	HFV-----NKTLLAAVPLAIDVLYVNSQGRGECFQRITITENAFKLVPLGTAAGCT	395
Db	282	NFITSQSTABERNELLS-----SCLCTLYTPD--EHFGIVPLEMAAYKPYIACNSGGP	333
Oy	396	TEIYLDGSGG-LTHPAGKEGAPLAATVRLASHAEORVSMWCKGYRVYEMEMEHNAE	454
Db	334	VEYKNGCVYGYLCEPTPEPDSSAMA-----RTENPELARMADEARHNVESVYKTCQ	389
Oy	455	RIAAVLKDVIRKSQE 469	
Db	390	KLNQYLVADVVSSPKE 404	
RESULT 10			
AAAG43120			
ID	AAG43120	standard, Protein, 397 AA.	
XX	AAAG43120;		
AC			
XX			
DT	18-OCT-2000	(first entry)	
XX			
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 53858.	
XX			
KW	protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		

[illegible]

Oy 459 VLKDLRKSGE 469
 Db 386 YLVNVVSPKE 396

RESULT 11
 AAG24139
 ID AAG24139 standard; Protein; 403 AA.

AC AAG24139;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27699.

KM Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134470.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137774.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144088.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144332.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144633.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147202.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148341.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158023.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158368.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159784.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 6.5%; Score 157; DB 21; Length 403;
Best Local Similarity 22.2%; Pred. No. 4.2e-07;
Matches 96; Conservative 74; Mismatches 187; Indels 76; Gaps 17;

QY 76 SKVLIVSH-ELSLSGPFLIMELAFLLRHVGSQV-VWITNORSOETNDVYSLEHRLN 133
DB 7 SKKNLAIITHPDIGIGAEERLIVDAAVELASHGHVHIFTSHHDKRCSEETLSGIQVIV 66
QY 134 HGVQVLPAR--GGEAVDIALKADLVITANTVAACKMLDPVLCDHPKPLP-----KI 182

DB 67 YG-SFLPHIFRYLHAXVAYLACLFLVALCVLGMSSFDVVLADQVSVVPLLTKRSSKV 125
QY 183 LMMIH-----EMRGHYFK-VEYVXHLDPVAGAMI--DSHTTAEYWSNRTSDRLK 228
DB 126 VFPCRFPDLLAKRTTTLRMWRKPIDFIEQOTTGMADILVNSFTASTF-ANTFKRLN 184
QY 229 IOWPQTVVHLGNSKELMEVADNVARVRLREHIFRESGLVREEDLFAIINSVSRGGQD 288
DB 185 AGGSRPALVYA-----VNIQDFIPHYKLNLSINREERKIND 225
QY 289 LPLQFYDALQLOHEKLVPRHIAVVG--SDVNAQTKFETQDLPVVKNTIHDRVH 345
DB 226 LNVSAF---AIIQCKKONLSDVLTITVAGGYDERLKENVEYLEEIRLSIAKEGSDRVNF 281
QY 346 V-----NKTLLAVA PYLAIDVLVONSQGRGECFRTTEIEMAFKLPVLTGAAGTTE 397
DB 282 ITSCSTAEKRLLS-----SCGLVLYPTD---EHFGIVPLEMAAKVPVIAQNSGQVPE 333
QY 398 IYLDGSLG-LIHPAGEGVAPLAKNIVRLASHAEQVSKGEKIGRYKEMFEMHMAERI 456
DB 334 TVKNGVTGYLCEPTPEDFSSAMA---FTEINPELANRKGAEARNHVESFSVKTFGQKL 389
QY 457 AAALKDVLKRSOE 469
DB 390 NQVLVDVVSPEKE 402

RESULT 12

AA624140
ID AA624140 standard; Protein; 395 AA.

AC AA624140;
XX

DT 17-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27700.
XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

OS Arabidopsis thaliana.
XX

PN EP1033405-A2.
XX

PD 06-SEP-2000.
XX

PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
XX

PR 05-MAR-1999; 99US-0123180.
XX

PR 09-MAR-1999; 99US-0123548
XX

PR 23-MAR-1999; 99US-0125788.
XX

PR 25-MAR-1999; 99US-0126264.
XX

PR 29-MAR-1999; 99US-0126785.
XX

PR 01-APR-1999; 99US-0127462.
XX

PR 06-APR-1999; 99US-0128234.
XX

PR 16-APR-1999; 99US-0129845.
XX

PR 19-APR-1999; 99US-0130077.
XX

PR 21-APR-1999; 99US-0130449.
XX

PR 23-APR-1999; 99US-0130510.
XX

PR 23-APR-1999; 99US-0130691.
XX

PR 28-APR-1999; 99US-0131449.
XX

PR 30-APR-1999; 99US-0132048.
XX

PR 30-APR-1999; 99US-0132407.
XX

PR 05-MAY-1999; 99US-0132484.
XX

PR 05-MAY-1999; 99US-0132485.
XX

PR 06-MAY-1999; 99US-0132486.
XX

PR 07-MAY-1999; 99US-0132487.
XX

PR 11-MAY-1999; 99US-0132488.
XX

PR 14-MAY-1999; 99US-0134218.
XX

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141827.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144452.
PR 20-JUL-1999; 99US-0144844.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156589.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158299.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161932.
PR 29-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 6.4%; Score 153.5; DB 21; Length 395;
Best Local Similarity 21.7%; Pred. No. 9,3e-07;

Matches 93; Conservative 74; Mismatches 187; Indels 75; Gaps 16;

QY 79 VLVSHLSLSGGLLMELAFLLRHVGSQV-WITNORSQETNDVYSLERHMLNMQV 137
DB 3 IAIHFDLIGIGAEHLIVDAVELASHGKHVIFSHHDSRCFETLSGIFQVTVYG-S 61
QY 138 VLPAR---GQEAVIDALKADLVILNTAVAGKMLDPVLDKDHVPKVL-----KILMWI 186
DB 62 FLFRHIFVRLHAXVAVYRCFLFVALCVLIGMSSFDVLADQSVVPELKLKSSKVVFC 121
QY 187 H-----EMRGHYFK-VVEYKHLPEVAGMT--DSHTTAEYMSRTSDCLKIQMP 232
DB 122 HFDLLLAHNTTLTTRMYRKPIDFIEBQTTGMADMIIVNSNFTASTF-ANFPKELANQGS 180
QY 233 QTVVHLGNSKELMEVAEDNVARVLRREHRESLGVSIEDLPALINSVSRGKQDLPFC 292
DB 181 RPAVLVPA-----VVIDGFIEPTTYKLTLSINPEKKKIIDLANS 221
QY 293 AFYQALQIQLHEKLVPRHIAVVG---SDVNAQTFEQLRDFVKNQTHDRVHEV--- 346
DB 222 AF---AIIICKKQNTSDVTLTVAGGYDERLKENVEYELBELRLAKGKGVSDRVMFTSC 277
QY 347 -----NKTLAAPYLAIDVLVONSQGECEFCGRITTEAFAFLPYLGAAGTTEIVD 401
DB 278 STABRELLS-----SCLCVLYPTD--EHGIVPLEMAAKKVIACNSGGFVEYVK 329
QY 402 GSGG-LHPAGKEGVAFLAKNIYRLASHAQRVSNGEKYGRVKEFMHMAERLAVL 460
DB 330 GVTGYLCEPTPEDFSMAA---RPLENELANMGAEARNHVVSFSVKTGCKNOYL 385
QY 461 KDLARKSOE 469
DB 386 VDVVSSPK 394

RESULT 13

AA93231
ID AA93231 standard; Protein; 368 AA.

AA93231;

04-SEP-2000 (first entry)

DE An Escherichia coli virulence protein.

KW Virulence protein; catA, catB, catC, tateC, mdgG, creC, recG, y99N;

KW eck1, iroD, iroC, iroE, mtd2, msl1; vaccine; infection;

KW Gram negative bacterium.

OS Escherichia coli.

XX MO200028038-A2.

XX 18-MAY-2000.

XX 09-NOV-1999; 99WO-GB03721.

XX 09-NOV-1998; 98GB-0024569.

XX 09-NOV-1998; 98GB-0024570.

XX 17-DEC-1998; 98GB-0027814.

PR 17-DEC-1998; 98GB-0027815.
PR 17-DEC-1998; 98GB-0027816.
PR 17-DEC-1998; 98GB-0027818.
PR 13-JAN-1999; 99GB-0000708.
PR 13-JAN-1999; 99GB-0000710.
PR 13-JAN-1999; 99GB-0000711.
PR 28-JAN-1999; 99GB-0001915.

PA (MICR-) MICROSCIENCE LTD.

PI Crooke HR, Clarke EE, Everest PH, Dougan G, Holden DW, Shea JE;

PI Feldman RG;

DR WPI; 2000-376550/32.

PT N-PSDB; AAA15182.

PT Peptide encoded by an operon including genes from Escherichia coli for

PT screening potential drugs, detecting virulence and treating conditions

PT associated with infection by a Gram negative bacterium -

PS Claim 2; Page 66-67; 122pp; English.

CC The present sequence represents an Escherichia coli virulence protein.

CC The specification describes virulence proteins which are encoded

CC by an operon including catA, catB, catC, tateC, mdgG, creC, recG, y99N,

CC eck1, iroD, iroC, iroE, mtd2 or msl-16 genes obtained from Escherichia

CC coli K1. The virulence proteins and polynucleotides, and their vaccines

CC are useful for screening potential drugs, for the detection of virulence,

CC and for treating or preventing conditions associated with infection by

CC a Gram negative bacterium particularly Escherichia coli.

CC Sequence 368 AA;

SQ

Query Match 6.4%; Score 152.5; DB 21; Length 368;
Best Local Similarity 21.7%; Pred. No. 1.1e-06;

Matches 90; Conservative 68; Mismatches 150; Indels 107; Gaps 19;

QY 65 VASGSPPLGPMKSKVLVLSHLSLSGGLLMELAFLLRHVGSQVWITNORSQETNDVT 124
DB 6 VITGIGLIG-GEKQVOLLADKLSGHHVKIISLGM-----SNKRV 46
QY 125 YSLERHMLNMGVOVLPARGQEAVIDALKADLVILNTAVAGKMLDPVLDKDHVPKILM 184
DB 47 PPSE-----NNVNVIVVNSKNISGVIKGCVR-----RDVIANFPKDI-- 85
QY 185 WHEMRGHYKVEYVHLPVA-----GAMIDSHTTAE--YMSRT-----SD----- 225
DB 86 ---VSHMPFRANIITRLSVIGIKNRPGIISTAHNNKNGGFRMLTYRTDCLSDCTNV 141
QY 226 -----RLKIQMPQTVVHLGNSKELMEVAEDNVARVLRREHRESLGVSIEDLFA 276
DB 142 SKEAVDEFRLKAFNP-AKAITMYNGIDTNKPFQDLARR-----EINDGINIKNDIILL 196
QY 277, LINSVSRGKQDPLQAFYQALQIQLHEKLVPRHIAVVGSDVNAQTFEQLD----- 332
DB 197 AAGRLLAKQYPLVLNAM---TLRPEHFKL-----IITG-----DGLBDEINM 237
QY 333 FVVKNTIHDRVHFVNTKLAAPYLAIDVLVONSQGECEFCGRITTEAFAFLPYLGAAGT 392
DB 238 LKKIKQLSNVSLGKVKNIAPFSACDIPVLS--RWEGLVVAEAMSCRIIVGIDS 295
QY 393 GGTTEIVLDSGTLLHPAGKEGVAFLAKNIYRLASHAQRVSNGEKYGRVKEFMHMAER 447
DB 296 GGIVREVI--GDDDFLVPISDS--TQLASKIEKL-SLSQIRHIGFRREIRILKNF 345

RESULT 14

ABB52953
ID ABB52953 standard; Protein; 368 AA.

XX ABB52953;

XX 11-FEB-2002 (first entry)

```

Oy      238 LKIKQLSNRSLSLGVYKXNIAIPYSAQDIFVLSS--RMEGFGLVAAEAMSCERLVVGTDG 295
Db      293 GGTTEIVLDGDSGTGLHPACKGCVAPLAKNIIVTASHAEQVYSGEKGVGXKMEF 447
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      296 GGQREVI--GGDDFLVPLSDS--TQLASKIEKL-SLSQIRDHIFGNNRERILKNF 345

RESULT 15
ID      AB852987
AB852987
ID      AB852987 standard; Protein; 368 AA.

XX      AC      AB852987;
XX      DT      11-FEB-2002 (first entry)
XX      DE      Escherichia coli polypeptide SEQ ID NO 1322.
XX      KM      Escherichia coli: B2/D+A-; antinflammatory; antibacterial;
XX      KM      immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX      KM      systemic infection; non-diarrhoeal infection; septicaemia;
XX      KM      pyelonephritis; antibiotic resistance.
XX      OS      Escherichia coli.
XX      PN      WO200166572-A2.
XX      PD      13-SEP-2001.
XX      PF      12-MAR-2001; 2001WO-EP03445.
XX      PR      10-MAR-2000; 2000FR-0003145.
XX      PR      02-FEB-2001; 2001FR-0001449.
XX      PA      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX      PI      Bingen E, Bonaccorsi S, Clermont O, Nassef X, Tinsley C;
XX      PI      WPI; 2001-550253/61.
XX      PT      A library of DNA fragments of Escherichia coli strains for the
XX      PT      phylogenic determination of a given strain comprises polynucleotides of
XX      PT      nature B2/D+ A -
XX      PS      Example 6; Fig 6; 646pp; English.
XX      CC      The invention relates to a library of DNA fragments of Escherichia coli
XX      CC      strains comprising polynucleotides (AB88677-AB88672 and AB88933)
XX      CC      and encoded proteins (AB852459-AB852919 and AB852954-AB853094) of nature
XX      CC      B2/D+A-. The polynucleotides have potential antinflammatory,
XX      CC      antibacterial, and immunosuppressive activity as part of pharmaceutical
XX      CC      compositions used to treat, palliate or prevent extra-intestinal B. coli
XX      CC      infections. The polypeptides are useful for determining the phylogenic
XX      CC      group of a given B. coli strain. These polypeptides can detect and treat
XX      CC      infection that include systemic and non-diarrhoeal infections such as
XX      CC      septicaemia, pyelonephritis and meningitis this is particularly
XX      CC      advantageous as bacterial resistance is increasing with the more
XX      CC      frequent use of broad spectrum antibiotics.
XX      CC      Sequence 368 AA;
XX      SQ
Oy      Query Match 6.4%; Score 152.5; DB 22; Length 368;
Oy      Best local similarity 21.7%; Pred. No. 1 ie-06;
Oy      Matches 90; Conservative 66; Mismatches 150; Indels 107; Gaps 19
Oy      65 VASGSPGFMRSKVLIVLSHSLSGGFLLMELATLRLHNGSQVWITNQSQETNDVT 124
Oy      6 VITELGIG-GAEKQVCLLAKRLKSLSGHHVIVTILGM-----SNKKV 46
Oy      125 YSLHEHRLHNGVQVLPARGQEAVDIALKADVLIVLNTAVAGKMLDPLKDHVKVPLKILW 184
Oy      47 FPSP-----NNNNVIVNNMSKNISGVIVGCRPL-----RDVIANFKPDI-- 85

```

QY 185 WIHEMCHYFKVYVYKHL.PFVA-----GAMIDSHTTAE--YVNSRT-----SD----- 225
Db 86 ----VISHMFHANIITRLSVIGIKNPGIISTAHKXEGYFRLTYRITDCLSDCCTNV 141
QY 226 -----RLKIOMPQTYVVHLGNSKEIMEVAEDNVARVLRBHIRESLGVSEDLIFA 276
Db 142 SKEAVDEFLRIKAFNP-AKAITWYNGIDTKFKFDLLAR-----EIRDGINIKNDLILL 196
QY 277 IINSVSRKGODLFLQAFYCALCLIOHEKLVPRIHAVVQSDVNAQTKEFOLRD----- 332
Db 197 AAGRLTLAKDYPNLNLNAM--TLLEPHFKL-----IIIC-----DGLRDEINM 237
QY 333 FVVKNTHDRVHFVNKTI LAVAPYLAIDVLVONSQGECEGGRITTEAMAFKLPVLTAA 392
Db 238 LIRKLOLSNRVSLGLGVKNKIAPYPSACDIFVLSS--RWEGFGLVVAEAMSCERIIVGTDS 295
QY 393 GGTTEIVLDGSTGLHPAGKEGVAPLAKNIVRLASHAEORVSMGEKGYGVKEMF 447
Db 296 GGVREVI--GDDDFLVPISSD--TQLASKIERL-SLQIRDHICFPNRERLIKNF 345

Search completed: June 30, 2003, 16:26:51
Job time : 74 secs

Db 132 -----KLIERSLFLTDKFLISESEYILANHISFNKSFSLINNGV--EVITGDSRN 182
Qy 261 HIRESLGVSEDLFLAINGSVSGKODLFLQAFYQALQLOHEKUKPRHIAVYVGSV 320
Db 183 EIEEIF--PNEDFLICWGRUSPPKEFFFIIDPAKLLQIRNDTNF-----IIVGCG- 232
Qy 321 NAOTFETOLDPVVKNITIDRVHFVNKTILAVAPYLAIDVLVONSQGRGECFGRIT-E 379
Db 233 ----ELSEIERMILNDGLGDKIYITGWDNPNRYIEKFDQALIFSRWEGU---SLTIAE 285
Qy 380 AAAPFLPYLGTAGGTEIYLDGSTGLHPAGKEGVAPLAKNIYRLASHAEQVSGEKG 439
Db 286 YMSQKKTILATNIGINDLITDGETGMLEIVGDLNSA-VSKSF-ELRNKKEVSNQLANNA 343
Qy 440 YGRVKEPF-MEHMAE 454
Db 344 YNKVVEQFSIEKQMAE 359

RESULT 2

US-08-746-682A-8
Sequence 8, Application US/08746682A
Patent No. 5766184
GENERAL INFORMATION:
APPLICANT: STINGELE, Francesca
APPLICANT: MOLLET, Beat
TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
TITLE OF INVENTION: EXOPOLYSACCHARIDES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americans
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746, 682A
FILING DATE: 14-NOV-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/597, 236
FILING DATE: 20-JUN-1995
APPLICATION NUMBER: EP 9520169.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Panucci A., Allen
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-746-682A-8

Query Match 5.2%; Score 125.5; DB 1; Length 372;
Best Local Similarity 21.8%; Pred. No. 0.00013;

Matches 69; Conservative 62; Mismatches 134; Indels 51; Gaps 17;

Qy 151 LKADVLITNTAVAGKLDVLPKDHVPKV-LP-KILMWHKRGHYRVE--YVKLPLPVA 206
Db 83 INPIVLVHSTFAG-----VGRIASIGLPTVYVNPB---GSPFMDNSYTKKLIF-- 131
Qy 207 GAMDSHTAEYWNSTSDRIKIMPTQYV--HLGNSKELMEVAEDNVARAVL-----RE 260
Db 132 -----KLIERSLFLTDKFLISESEYILANHISFNKSFSLINNGV--EVITGDSRN 182

Qy 261 HIRESLGVSEDLFLAINGSVSGKODLFLQAFYQALQLOHEKUKPRHIAVYVGSV 320
Db 183 EIEEIF--PNEDFLICWGRUSPPKEFFFIIDPAKLLQIRNDTNF-----IIVGCG- 232
Qy 321 NAOTFETOLDPVVKNITIDRVHFVNKTILAVAPYLAIDVLVONSQGRGECFGRIT-E 379
Db 233 ----ELSEIERMILNDGLGDKIYITGWDNPNRYIEKFDQALIFSRWEGU---SLTIAE 285
Qy 380 AAAPFLPYLGTAGGTEIYLDGSTGLHPAGKEGVAPLAKNIYRLASHAEQVSGEKG 439
Db 286 YMSQKKTILATNIGINDLITDGETGMLEIVGDLNSA-VSKSF-ELRNKKEVSNQLANNA 343
Qy 440 YGRVKEPF-MEHMAE 454
Db 344 YNKVVEQFSIEKQMAE 359

RESULT 3

US-09-134-001C-3206
Sequence 3206, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134, 001C
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064, 964
PRIOR FILING DATE: 1997-08-14
PRIOR APPLICATION NUMBER: US 60/055, 779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3206
LENGTH: 387
TYPE: PRN
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3206

Query Match 5.0%; Score 119; DB 4; Length 387;
Best Local Similarity 18.7%; Pred. No. 0.00069;

Matches 80; Conservative 77; Mismatches 173; Indels 98; Gaps 15;

Qy 87 SLSSGFLIMELAFILRHVSQVYVWITNQSGETNDVYSLERHMLN---HGYOVLPRAG 143
Db 17 SMGSGIATLGLTGMARGHEVHPT-----TSNIPRIKRPDENMTFOVEV----N 65
Qy 144 QEAV-----DIALKADVLITNTAVAGKLDVLPKDHVPKVLKILMWHK 188
Db 66 QYAVFOYPPYDITLSTKLSVDIOEYDLIMHYA-----VPHAVCGIL--AKQ 112
Qy 189 MRGHYFVEYVGLPFGVAGMID-----SHTLAEVWNSRSTDRK 228
Db 113 MSGGNKIMTTHGHDITLVGYDHTLONAIKFGIEQSDIYTSVSHLQ----- 161
Qy 229 IQMPQTYVHLGNSKELMEV--AEDNVARAVLEHIRESLGVSEDLFLAINGSVSGK 285
Db 162 ----QYVEI--INTKKEIIPYVFNRENEPPTRNHEELKDCYGISPEKKVLHVSFPAKY 216
Qy 286 GODLFLQAFYQALQLOHEKUKPRHIAVYVGSVDNAQTFTOLDPVVKNITIDRVHF 345
Db 217 RIDVLVETFAKY-----HESIP-----SKLILGCGPELIDMRKARELDET-----HVL 263
Qy 346 VNKTLAVAPYLAIDVLVONSQGRGECFGRITIAVAFKLPVLGTAGGTEIYLDGSTG 405
Db 264 LGRQNDVSAFYQLSDVLVLLSE--KESFELTLEAMKTVLPISGHAQIGKEVIRHEETG 321
Qy 406 LHPAGKEGVAPLAKNIYRLASHAEQVSMGEKGYGVKEMPEHMAERLAVALKDVLR 465
Db 322 FIVYDIDSTQA--AKAIVKLSNPELYQKMSQMLKDIARPSDILIDYVENYRKMLE 379
Qy 466 KSDHRS 473

Db 380 OCEANNE 387

RESULT 4
US-09-403-768-8
Sequence 8, Application US/09403768

Patent No. 6444804
GENERAL INFORMATION:

APPLICANT: Lam Joseph S.
APPLICANT: De Kievit, Tereza R.

APPLICANT: Burrows, Lori L.

APPLICANT: Walsh, Andrew

TITLE OF INVENTION: No. 6444804e1 Proteins Involved in the Synthesis and Assembly

TITLE OF INVENTION: of Core Lipopolysaccharide of Pseudomonas Aeruginosa

FILE REFERENCE: 6580-177

CURRENT APPLICATION NUMBER: US/09/403,768

PRIOR APPLICATION NUMBER: U.S. 60/045,418

PRIOR FILING DATE: 1997-05-02

PRIOR APPLICATION NUMBER: U.S. 60/046,149

PRIOR FILING DATE: 1997-05-09

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 373

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa RFag (Maag)

FEATURE: Description of Organism: Pseudomonas aeruginosa

OTHER INFORMATION: serotype O5 strain PAOI

US-09-403-768-8

Query Match 4.5%; Score 108.5; DB 4; Length 373;

Best Local Similarity 20.9%; Pred. No. 0.0083;

Matches 46; Conservative 39; Mismatches 98; Indels 45; Gaps 8;

Db 228 KIQMPQYVYVHLGNSKELMEVADNVA--RV-----LREHRESLGRSDDLPAIIN 219

145 EVQOP-LLVKHYGTQAEHFHLPFGISODRAPANADVAERFRREBGLBEDDLLVQIG 203

280 SVSRGKGDLPLOAFYQALQIHEKLVPRHIAVVGSDVNAQTFFETQLDEPVNNT- 318

204 SGRTKGLDRSLKALSALPKALRR-----RRLAIQGD-----DPKFIQIAA 248

Qy 339 --IHDRVHFVNKTLAVPYLAIDVLYONSQGRGECFRRTITAMAFKLPVIGTRAGTT 396

249 LGANDQVQILKRS DIPFLIGADLIHPAY--NNTGTVLLELVSGLPVLTVDVCGYA 306

Qy 397 EIVDGSSTGLHPAGKE-----GVAFLAKIVYLAASHAE 430

307 HYIEMADAGRVLPSPFFQDSINRLLMLEMDAPAPAAASRGGLAVDHD 356

RESULT 5
US-08-941-445A-5
Sequence 5, Application US/08941445A

Patent No. 6107060
GENERAL INFORMATION:

APPLICANT: Keeling, Peter

APPLICANT: Guan, Hanning

TITLE OF INVENTION: Starch Encapsulation

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee, Wimmer and Sullivan, P.C.

STREET: 5370 Manhattan Circle

CITY: Boulder

STATE: CO

COUNTRY: US

ZIP: 80303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/941,445A

FILING DATE: 30-SEP-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/026,855

FILING DATE: 30-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Wimmer, Ellen P

REGISTRATION NUMBER: 28,547

REFERENCE/DOCKET NUMBER: 89-97

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 533 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-941-445A-5

Query Match 4.5%; Score 108.5; DB 3; Length 533;

Best Local Similarity 22.8%; Pred. No. 0.015;

Matches 44; Conservative 35; Mismatches 73; Indels 41; Gaps 7;

Db 254 ARRVIREHIRESG--VRESDLLPAIINSVSRGKGDLPLOAFYQALQIHEKLVPR 311

302 AKLKNKELQEVGLPYDNIPLVAFIORLEBOKPDPVAAALIPLMEMVED-----V 354

Qy 312 HAVVVGSDVNAQTFFETQLDFVFNKTHDRVHFNKTLAVPYLAAL-----DIVV 363

355 QIVLLGT--GKKKFERMLMSAEK-----FPKVRVAVKFAALAHHIMAGADVIA 403

Qy 364 QNSQGRGECFRRTITAMAFKLPVIGTRAGTTIIVDSTG-----LHPAGK 412

404 VTS--RFPCLQIQDQMRGTPCACAATGSLVDVTIIEKTFHNGRLSDVCNVERADV 461

Qy 413 EGVAFLAKIVYRL 425

Db 462 KKVATLQRAIKV 474

RESULT 6
US-08-861-464-10
Sequence 10, Application US/08861464

Patent No. 5874210
GENERAL INFORMATION:

APPLICANT: Guarente, Leonard P.

APPLICANT: Austriaco Jr., Nicanor

TITLE OF INVENTION: Genes Determining Cellular Senescence

TITLE OF INVENTION: in Yeast

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/861,464

FILING DATE: 22-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/396,001
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09351
; FILING DATE: 15-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,408
; FILING DATE: 16-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-861-464-10

```

```

Query Match      4.4%; Score 105.5; DB 2; Length 508;
Best Local Similarity 21.4%; Pred. No. 0.029;
Matches 97; Conservative 75; Mismatches 140; Indels 141; Gaps 26;

```

```

QY 101 LIRHVGSOV-----WITNORSQETNDV---TYSL-----EHRMLNHGVQLPARQOE 145
DB 105 MLRHAESAIVEAYVNDKAILBQRNMLTEELYGNTFOLYSADHRLTDLKYLEVQPEKIEL 164
QY 146 AVD-----IALKADLVILNTAVAGKMLD-----PYLKHVPKVLPRILMWHMERGH 192
DB 165 IMDEMKQILTPMAOK-EAVIKHSLVHKYFLDPFTYAP-----PKLRSMTEAIRF----- 213
QY 193 YFKVEYVHLRPFVAGAMIDSHTTAAYNSTRSDR-LKIQMPQTYVYHLSGNSK--ELMEVA 249
DB 214 --AVVYLAHTH--DGARVAMHC---LMHGTPEKDKRVIVKTMKYVEKVAHQYSHLVULA 266
QY 250 -----EDNVARVILREHIRESLGVRSEDLFPALINSVSRGQDLFLQAFYQALQILQH 303
DB 267 AFDICIDTGLVKQIILSEIISLP-----SIYND-KYGRKVLVLSPRDPHATVR- 316
QY 304 EKLKVRIRHAAVVGSDVNAQTKEFTQLRDPVVKNTIHDRVHPNKTIAVAPYLA--AIDV 361
DB 317 -----EILEVLQKDGDAVSHSKDTEVRRELLSEI-----SPALLSYLOEHAQEV 361
QY 362 LVNOSGREGCEFGRTITTEMAFKLPVLGTAAAGT-----TEIYLDGSTGLHPAGKEGVAP 417
DB 362 VLDSK-----ACVLVSDILGSATGVQPTMNAIASLATG-LHFGKDGDELH 407
QY 418 LAKNIVRLASH-----AEQVSMGEKG-YGRVKEMFMEH----- 450
DB 408 IAH--PAGHLVLKMLIEQDKMKENGREGCFPAKTLVHEIVGKMKLKSMAVNRGAITLS 464
QY 451 -----HMAERIAAVLKDV---LRKSOEHSR 472
DB 465 SLIASCDELVANRKYKALKSLIPTLEKTKSTSK 497

```

```

RESULT 7
US-08-396-001-10
; Sequence 10, Application US/08396001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence in
; TITLE OF INVENTION: Yeast
; NUMBER OF SEQUENCES: 16

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-396-001-10

```

```

Query Match      4.4%; Score 105.5; DB 2; Length 508;
Best Local Similarity 21.4%; Pred. No. 0.029;
Matches 97; Conservative 75; Mismatches 140; Indels 141; Gaps 26;

```

```

QY 101 LIRHVGSOV-----WITNORSQETNDV---TYSL-----EHRMLNHGVQLPARQOE 145
DB 105 MLRHAESAIVEAYVNDKAILBQRNMLTEELYGNTFOLYSADHRLTDLKYLEVQPEKIEL 164
QY 146 AVD-----IALKADLVILNTAVAGKMLD-----PYLKHVPKVLPRILMWHMERGH 192
DB 165 IMDEMKQILTPMAOK-EAVIKHSLVHKYFLDPFTYAP-----PKLRSMTEAIRF----- 213
QY 193 YFKVEYVHLRPFVAGAMIDSHTTAAYNSTRSDR-LKIQMPQTYVYHLSGNSK--ELMEVA 249
DB 214 --AVVYLAHTH--DGARVAMHC---LMHGTPEKDKRVIVKTMKYVEKVAHQYSHLVULA 266
QY 250 -----EDNVARVILREHIRESLGVRSEDLFPALINSVSRGQDLFLQAFYQALQILQH 303
DB 267 AFDICIDTGLVKQIILSEIISLP-----SIYND-KYGRKVLVLSPRDPHATVR- 316
QY 304 EKLKVRIRHAAVVGSDVNAQTKEFTQLRDPVVKNTIHDRVHPNKTIAVAPYLA--AIDV 361
DB 317 -----EILEVLQKDGDAVSHSKDTEVRRELLSEI-----SPALLSYLOEHAQEV 361
QY 362 LVNOSGREGCEFGRTITTEMAFKLPVLGTAAAGT-----TEIYLDGSTGLHPAGKEGVAP 417
DB 362 VLDSK-----ACVLVSDILGSATGVQPTMNAIASLATG-LHFGKDGDELH 407
QY 418 LAKNIVRLASH-----AEQVSMGEKG-YGRVKEMFMEH----- 450
DB 408 IAH--PAGHLVLKMLIEQDKMKENGREGCFPAKTLVHEIVGKMKLKSMAVNRGAITLS 464
QY 451 -----HMAERIAAVLKDV---LRKSOEHSR 472
DB 465 SLIASCDELVANRKYKALKSLIPTLEKTKSTSK 497

```

```

RESULT 8
US-09-323-433A-10
; Sequence 10, Application US/09323433A
; Patent No. 6218512
; GENERAL INFORMATION:

```

Query Match	4.4%;	Score 105.5;	DB 4;	Length 508;
Best Local Similarity	21.4%;	Pred. No. 0.029;		
Matches	97;	Conservative	75;	Mismatches 140;
				Indels 141;
				Gaps 26;

```

, RESULT 9
, US-09-335-409-5
, Sequence 5, Application US/09335409
, Patent No. 6121029
, GENERAL INFORMATION:
, APPLICANT: Schupp, Thomas
, APPLICANT: Ligon, James
, APPLICANT: Molnar, Istvan
, APPLICANT: Zirkle, Ross
, APPLICANT: Cyr, Devon
, APPLICANT: Goerlach, Joern
, TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

```

Query Match	4.4%;	Score 104.5;	DB 3;	Length 7257;
Best Local Similarity	19.7%;	Pred. No. 3.2;		
Matches 100; Conservative	61;	Mismatches 146;	Indels 201;	Gaps 21;

```

RESULT 10
US-09-568-102-5
; Sequence 5, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligou, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Rose
; APPLICANT: Cyr, Devon
; APPLICANT: Goelisch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257

```

```

; TYPE: PROT
; ORGANISM: Sorangium cellulosum
US-09-568-102-5

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21

OY 58 PALNTRAVASGSPFLGFRSKLVLVVSHSLSGSPLLMELAF-----LIRHVGSO 108
DB 6623 PALTLAQAALPVPFTMTAYGLV-HIGRLRGERVLIHSATGCTGLAAVQIARHGAE 6686
OY 109 V-----VITNQSGSEINDVTYSLEHRLNNGVQLVPARGDAVDIALAKADLVII 158
DB 6661 IFPATGPEKRAMLEQ-----GIAHMQDSRLDFAQVLAATKEGVD-----VVL 6722
OY 159 NTAVAGKMDLPVLKDHVPKVLKILMWIHMERGHYFKVEYVKLPPVAGAMIDSHTAEX 218
DB 6728 NS-----LSGAALDA----- 6733
OY 219 MNSRTDRLKIQMPQTYVHVLNLSYELMEVAEDNNAARYL-REHIRESGRSED----- 272
DB 6738 -----SLSTIVPRGFIELGK-----DIYADRSGLAHFKKSLSYSAVDLAGLA 6782
OY 273 -----LLFAINSVSRGKGODLFLOAF-----YQALQIQH----- 303
DB 6783 VRSPERVAALLAEVVDLIRAGALQPLPVEIFFLSRAADAFFKQAQNHGKVLVALEDD 6842
OY 304 EKKVPRIHAAVYVSDPNAGTFFEQDLRDVYVNTIHRVHFNKTLAAVPLA----- 357
DB 6843 VRIIRVSGEVAITAD-----GAYLVGGVG-----GLGLSYAGMLAEQGAGH 6885
OY 358 -----AIDVLVONSQSGEGCFG-RITF-----EMAAFKPLVLTG-- 390
DB 6886 LVLVGRSGAVAEQOTVAVALTAHAGARVTVARADVAQRMERILREVTASGMPLRGVH 6945
OY 391 AAGGTTEIVLDGST-----GLHPAGKEGVAPLAKNYTRLASAEGRVSMGE 437
DB 6946 AAGIDDDGLMLQOQIPAFRFAWAPKVGALHLHALTREAPLS-FVLVASGAGLGSFQ 7004
OY 438 KGYGRVK--EMFEHHMARIADLVKD 462
DB 7005 GNYAANFTDLALHHRRAQGLPALSID 7032

RESULT 11
US-09-567-969-5
; Sequence 5, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PROT
; ORGANISM: Sorangium cellulosum
US-09-567-969-5

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

```

QY	58	PAALNTAVASGSPGFMRKSLVILVSHSLISGGPILIMEALF-----LIRHVSG	108
Db	6623	PAALTAQAAPALPAFAFTATAYGV--HGRRLRGERVILHATGCGGLAAVQIARHIGAE	6680
QY	109	V-----VITNRSQETNDVTSLSHRMLNHGVQLPARGOEVDIALKADVL	158
Db	6661	IFATAGTEPKRAMREQ-----GIAHWMSRLDPAEOVLAATKGEVD-----VVL	6727
QY	159	NTAVAGKMLDPLVDKHVPKILIMIHMRGHYFKEVYKHLPEVAGAMIDSHTTAEY	218
Db	6728	NS-----LSGAALDA-----	6733
QY	219	WNSRTDLKIQMPQTYVHHLGNSKELMEVAEDNNARVL-REHTIESIGVRSDD-----	272
Db	6738	-----SLSTLVPDGRFTLEKRT-----DIYADRSGLLHAFKRSISYSAVDLAGLA	6783
QY	273	-----LLFAINSVSRRKGDDPLFLQAF-----YQALQLIQH-----	303
Db	6783	VRRPERVALLAEVVDILRLARGALQPLPEVIFPLSPADAAPKQAQNHGKLVLALEDD	6842
QY	304	EKKVPRHIAVAVGSDPNMQTFEQLRDVVYKNTIHDRVHFNKTLVAAVYA-----	357
Db	6843	VRIKRVSGEGLAIAAD-----GAYLVTGGG-----GLGLSVAGMLBEGGAGH	6885
QY	358	-----AIDVLVNSGRCGECGF-RITL-----EMAEKPLPVLTG--	390
Db	6886	LVLVGRSGAVAEBOQTAAVVALEHAGARVTVAADVADRAQWERILRETTAASGMPLGCVH	6945
QY	391	AAAGTEIYLDGST-----GLLHPAKGEGVAPLAKNIYPLASIAEGRVSMGE	437
Db	6946	AAIGLDDGLMQQTPARFPAVMAKRVGALHIALHTREAPLS-FVLYASGAGILSSPQQ	7004
QY	438	KCGIKRYK--EMFEHMHMERIAVLKD	462
Db	7005	GNYAAANTFTDLALHRRAGGLPALSID	7032
RESULT 12			
US-09-568-480-5			
Sequence 5, Application US/09568480			
Patent No. 6355458			
GENERAL INFORMATION:			
APPLICANT: Schupp, Thomas			
APPLICANT: Ligon, James			
APPLICANT: Molnar, Istvan			
APPLICANT: Zirkle, Ross			
APPLICANT: Cyr, Devon			
APPLICANT: Goerlach, Joern			
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES			
FILE REFERENCE: 4-30582A			
CURRENT APPLICATION NUMBER: US/09/568,480			
CURRENT FILING DATE: 2000-05-10			
PRIOR APPLICATION NUMBER: 09/335,409			
PRIOR FILING DATE: 1999-06-17			
NUMBER OF SEQ ID NOS: 30			
SOFTWARE: Patencin Ver. 2.0			
SEQ ID NO 5			
LENGTH: 7257			
TYPE: PRT			
ORGANISM: Sorangium cellulosum			
US-09-568-480-5			
Query Match			
4.4%; Score 104.5; DB 4; Length 7257;			
Best Local Similarity 19.7%; Pred. No.3.2;			
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;			
QY	58	PAALNTAVASGSPGLFMRKSLVILVSHSLISGGPILIMEALF-----LIRHVSG	108
Db	6623	PAALTAQAAPALPAFAFTATAYGV--HGRRLRGERVILHATGCGGLAAVQIARHIGAE	6680
QY	109	V-----VITNRSQETNDVTSLSHRMLNHGVQLPARGOEVDIALKADVL	158
Db	6661	IFATAGTEPKRAMREQ-----GIAHWMSRLDPAEOVLAATKGEVD-----VVL	6727
QY	159	NTAVAGKMLDPLVDKHVPKILIMIHMRGHYFKEVYKHLPEVAGAMIDSHTTAEY	218
Db	6728	NS-----LSGAALDA-----	6733
QY	219	WNSRTDLKIQMPQTYVHHLGNSKELMEVAEDNNARVL-REHTIESIGVRSDD-----	272
Db	6738	-----SLSTLVPDGRFTLEKRT-----DIYADRSGLLHAFKRSISYSAVDLAGLA	6783
QY	273	-----LLFAINSVSRRKGDDPLFLQAF-----YQALQLIQH-----	303
Db	6783	VRRPERVALLAEVVDILRLARGALQPLPEVIFPLSPADAAPKQAQNHGKLVLALEDD	6842
QY	304	EKKVPRHIAVAVGSDPNMQTFEQLRDVVYKNTIHDRVHFNKTLVAAVYA-----	357
Db	6843	VRIKRVSGEGLAIAAD-----GAYLVTGGG-----GLGLSVAGMLBEGGAGH	6885
QY	358	-----AIDVLVNSGRCGECGF-RITL-----EMAEKPLPVLTG--	390
Db	6886	LVLVGRSGAVAEBOQTAAVVALEHAGARVTVAADVADRAQWERILRETTAASGMPLGCVH	6945
QY	391	AAAGTEIYLDGST-----GLLHPAKGEGVAPLAKNIYPLASIAEGRVSMGE	437
Db	6946	AAIGLDDGLMQQTPARFPAVMAKRVGALHIALHTREAPLS-FVLYASGAGILSSPQQ	7004
QY	438	KCGIKRYK--EMFEHMHMERIAVLKD	462
Db	7005	GNYAAANTFTDLALHRRAGGLPALSID	7032
US-09-568-480-5			
Sequence 5, Application US/09568480			
Patent No. 6355458			
GENERAL INFORMATION:			
APPLICANT: Schupp, Thomas			
APPLICANT: Ligon, James			
APPLICANT: Molnar, Istvan			
APPLICANT: Zirkle, Ross			
APPLICANT: Cyr, Devon			
APPLICANT: Goerlach, Joern			
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES			
FILE REFERENCE: 4-30582A			
CURRENT APPLICATION NUMBER: US/09/568,480			
CURRENT FILING DATE: 2000-05-10			
PRIOR APPLICATION NUMBER: 09/335,409			
PRIOR FILING DATE: 1999-06-17			
NUMBER OF SEQ ID NOS: 30			
SOFTWARE: Patencin Ver. 2.0			
SEQ ID NO 5			
LENGTH: 7257			
TYPE: PRT			
ORGANISM: Sorangium cellulosum			
US-09-568-480-5			
Query Match			
4.4%; Score 104.5; DB 4; Length 7257;			
Best Local Similarity 19.7%; Pred. No.3.2;			
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;			
QY	58	PAALNTAVASGSPGLFMRKSLVILVSHSLISGGPILIMEALF-----LIRHVSG	108
Db	6623	PAALTAQAAPALPAFAFTATAYGV--HGRRLRGERVILHATGCGGLAAVQIARHIGAE	6680
QY	109	V-----VITNRSQETNDVTSLSHRMLNHGVQLPARGOEVDIALKADVL	158
Db	6661	IFATAGTEPKRAMREQ-----GIAHWMSRLDPAEOVLAATKGEVD-----VVL	6727
QY	159	NTAVAGKMLDPLVDKHVPKILIMIHMRGHYFKEVYKHLPEVAGAMIDSHTTAEY	218
Db	6728	NS-----LSGAALDA-----	6733
QY	219	WNSRTDLKIQMPQTYVHHLGNSKELMEVAEDNNARVL-REHTIESIGVRSDD-----	272
Db	6738	-----SLSTLVPDGRFTLEKRT-----DIYADRSGLLHAFKRSISYSAVDLAGLA	6783
QY	273	-----LLFAINSVSRR	

Db 6681 IFATAGTEKRAMLRQ-----GIAHWDSRLDPAEQVLAATKGEVD-----VVL 6727
QY 159 NTAVAGKMLDVLKDHVPKVLPKILMWHMERGHYFVEVYVHLFPVAGMIDSHHTAAY 218
Db 6728 NS-----LSGAIDA----- 6737
QY 219 WNSRTSDRLKIOMPOQYVYVHLSKELMEVADNVARVL-REHIRESLGVRSD----- 272
Db 6738 -----SLSTLVPDGRFIELGKT-----DIYADRSIGLAHFRKSLSYSXVDLAGIA 6782
QY 273 -----LFPAINSVSRKGGODFLQAF-----YQALQLOIQR----- 303
Db 6783 VRPERVALLAEVVDLARGALQPLVEIFPLSRADAFRKAQOHLGTLVLEDDP 6842
QY 304 EKLKVPRIHAAVVGSDVNAQTKFETQLRDFVKNKTIHDRVHPVKNKTLAAVAYLA----- 357
Db 6843 VIRIVPGESGVALRAD-----GAYLVGTGLG-----GLGSLVAGWLAEOGAGH 6885
QY 358 -----AIDVLVONSQGRGECFG-RITI-----EMAFKFLPVLTG-- 390
Db 6886 LVLVGSGAVSAEQOTVAALLENHARGARTVVARADVADRAQMERILREVTASGMPLRGVH 6945
QY 391 AAGGTEIVLDGST-----GLHPAGEGVADPAKNIYRLASHAEOVRSMGE 437
Db 6946 AAGIIDDGLMOQTPARFRAVMAPKVRGALHLHALTREAPLS-FPVLVYASGAGLLSPQO 7004
QY 438 KGYGRVK---EMPEHMAERIAVAKD 462
Db 7005 GNYAANTFLDALHHRRAQGLPALSTD 7032

RESULT 13
US-09-568-486-5
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligot, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-5

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

QY 58 PALNTAVASGSPGLFMRSKVLVLVSHLSLGGPLLMELAF-----LIRHVSQ 108
Db 6623 PALTLAQAALPFAFTAYVGLV-HIGRLRAGERVLHSATGCTGLAAVQARHLGAE 6680
QY 109 V-----WITNQRSGEITNDVTYSLEHRLMHGCVQVLPARGQENVDAKADLYIL 158
Db 6681 IFATAGTEKRAMLRQ-----GIAHWDSRLDPAEQVLAATKGEVD-----VVL 6727
QY 159 NTAVAGKMLDVLKDHVPKVLPKILMWHMERGHYFVEVYVHLFPVAGMIDSHHTAAY 218
Db 6728 NS-----LSGAIDA----- 6737
QY 219 WNSRTSDRLKIOMPOQYVYVHLSKELMEVADNVARVL-REHIRESLGVRSD----- 272

Db 6738 -----SLSTLVPDGRFIELGKT-----DIYADRSIGLAHFRKSLSYSXVDLAGIA 6782
QY 273 -----LFPAINSVSRKGGODFLQAF-----YQALQLOIQR----- 303
Db 6783 VRPERVALLAEVVDLARGALQPLVEIFPLSRADAFRKAQOHLGTLVLEDDP 6842
QY 304 EKLKVPRIHAAVVGSDVNAQTKFETQLRDFVKNKTIHDRVHPVKNKTLAAVAYLA----- 357
Db 6843 VIRIVPGESGVALRAD-----GAYLVGTGLG-----GLGSLVAGWLAEOGAGH 6885
QY 358 -----AIDVLVONSQGRGECFG-RITI-----EMAFKFLPVLTG-- 390
Db 6886 LVLVGSGAVSAEQOTVAALLENHARGARTVVARADVADRAQMERILREVTASGMPLRGVH 6945
QY 391 AAGGTEIVLDGST-----GLHPAGEGVADPAKNIYRLASHAEOVRSMGE 437
Db 6946 AAGIIDDGLMOQTPARFRAVMAPKVRGALHLHALTREAPLS-FPVLVYASGAGLLSPQO 7004
QY 438 KGYGRVK---EMPEHMAERIAVAKD 462
Db 7005 GNYAANTFLDALHHRRAQGLPALSTD 7032

RESULT 14
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligot, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

QY 58 PALNTAVASGSPGLFMRSKVLVLVSHLSLGGPLLMELAF-----LIRHVSQ 108
Db 6623 PALTLAQAALPFAFTAYVGLV-HIGRLRAGERVLHSATGCTGLAAVQARHLGAE 6680
QY 109 V-----WITNQRSGEITNDVTYSLEHRLMHGCVQVLPARGQENVDAKADLYIL 158
Db 6681 IFATAGTEKRAMLRQ-----GIAHWDSRLDPAEQVLAATKGEVD-----VVL 6727
QY 159 NTAVAGKMLDVLKDHVPKVLPKILMWHMERGHYFVEVYVHLFPVAGMIDSHHTAAY 218
Db 6728 NS-----LSGAIDA----- 6737
QY 219 WNSRTSDRLKIOMPOQYVYVHLSKELMEVADNVARVL-REHIRESLGVRSD----- 272
Db 6738 -----SLSTLVPDGRFIELGKT-----DIYADRSIGLAHFRKSLSYSXVDLAGIA 6782
QY 273 -----LFPAINSVSRKGGODFLQAF-----YQALQLOIQR----- 303
Db 6783 VRPERVALLAEVVDLARGALQPLVEIFPLSRADAFRKAQOHLGTLVLEDDP 6842

```

QY 304 EKLKVPRIHAVVSGDVNAQKFEQTLRD FVKNKTIHDRVHFNKTLAVPYLA----- 357
DB 6843 VRIIPVBSGVAIRAD-----GAVLVYTGGLG-----GLGLSVGMWLAGAGAH 6885
QY 358 -----AIDVLVONSQGRGCFG-RTTI-----EMAFKLPVIGT-- 390
DB 6886 LVLVGSGAVSAEQOTIVAALEHAGARVAVRADVDADRAQWERILIREVTASGMPLEGGVH 6945
QY 391 AAGGTEIVLDGST-----GLHPAGKEGVAPLAKNIVRLASHAQVSMGE 437
DB 6946 AAGILDDGLMOQTPARFRAYVAPKVRGALHIALHATREAPLS-FVLVYASGAGLIGSPQ 7004
QY 438 KGYGRVK---EMFMEHHMAERIAAVLKD 462
DB 7005 GNYAANTFLDALHHRRAOGLPALSTD 7032

```

```

RESULT 15
US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-5

```

```

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

```

```

QY 58 PAALNTAVASGSPFGFMRSKTIVLVSHSLSCGPILLMELAF-----LHRVSGQ 108
DB 6623 PAALTAQAALPVAFTAWGLV--HLGRLRAGERVLIHSATGTGIAAVQIARHLSAE 6680
QY 109 V-----VMTNQSQETNDVYSLERMLNHGVQLPARGQAVDIALKADLVIL 158
DB 6681 IFATNGTPEKRAWLREQ-----GIAHVMSRLDPABQVLAATKGEVD-----VVL 6727
QY 159 NTAVAGKMLDPVLKDHVPKVLPKILMWIHEMRGHYFKVEYVYKHLPFVAGAMIDSHTAEY 218
DB 6728 NS-----LSGAALDA----- 6737
QY 219 WNSRTSRLKIQMPTVYVHLGNSKELMEVADENYARVL-REHIRESLGVNSD----- 272
DB 6738 -----SLSTLVDPDGRFELGKT-----DIYADSLGLAHFRKSLSYSVAVDLAGLA 6782
QY 273 -----LFPALINSVSRGCGDPLQAF-----YQALQLIQH----- 303
DB 6783 VRPERVAALLAEVVDLLARGLQLPVEIFPLSRADAQFRKMAQOHGKVLALBDPD 6842
QY 304 EKLKVPRIHAVVSGDVNAQKFEQTLRDVAVYKNTIHDRVHFNKTLAVPYLA----- 357
DB 6843 VRIIPVBSGVAIRAD-----GAVLVYTGGLG-----GLGLSVGMWLAGAGAH 6885
QY 358 -----AIDVLVONSQGRGCFG-RTTI-----EMAFKLPVIGT-- 390
DB 6886 LVLVGSGAVSAEQOTIVAALEHAGARVAVRADVDADRAQWERILIREVTASGMPLEGGVH 6945

```

```

QY 391 AAGGTEIVLDGST-----GLHPAGKEGVAPLAKNIVRLASHAQVSMGE 437
DB 6946 AAGILDDGLMOQTPARFRAYVAPKVRGALHIALHATREAPLS-FVLVYASGAGLIGSPQ 7004
QY 438 KGYGRVK---EMFMEHHMAERIAAVLKD 462
DB 7005 GNYAANTFLDALHHRRAOGLPALSTD 7032

```

```

Search completed: June 30, 2003, 16:31:00
Job time : 29 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2003, 16:29:51 ; Search time 51 Seconds
(without alignments)
1016.921 Million cell updates/sec

Title: US-09-938-294-45
Perfect score: 2397
Sequence: 1 MAKTPFAVAAGSGRPVH.....ERIAVLKVKRSGHSRS 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 424699 seqs, 109646833 residues
Total number of hits satisfying chosen parameters: 424699

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubppaa/PTC_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubppaa/CTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUB pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep.*
10: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep.*
11: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep.*
12: /cgn2_6/ptodata/2/pubppaa/US60_PUB pep.*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181	7.6	191	10	US-09-924-358-22
2	179	7.5	144	10	US-09-924-358-25
3	162.5	6.8	385	10	US-09-767-041-18
4	150	6.3	387	9	US-09-738-626-3890
5	134.5	5.6	416	10	US-09-924-358-5
6	128.5	5.4	409	9	US-09-738-626-4738
7	127	5.3	381	9	US-09-738-626-3951
8	126.5	5.3	416	10	US-09-815-028-5
9	118	4.3	371	9	US-09-870-759-107
10	113	4.7	720	9	US-10-217-700-14
11	111	4.6	323	9	US-10-174-590-296
12	111	4.6	323	9	US-10-176-758-296
13	111	4.6	323	9	US-10-175-737-296
14	111	4.6	323	9	US-10-173-706-296
15	111	4.6	323	9	US-10-175-738-296
16	111	4.6	323	9	US-10-175-752-296
17	111	4.6	323	9	US-10-176-482-296
18	111	4.6	323	9	US-10-176-482-296
19	111	4.6	323	9	US-10-176-757-296

20	111	4.6	323	9	US-10-176-913-296	Sequence 296, App
21	111	4.6	323	9	US-10-180-552-296	Sequence 296, App
22	111	4.6	323	9	US-10-180-557-296	Sequence 296, App
23	111	4.6	323	9	US-10-173-700-296	Sequence 296, App
24	111	4.6	323	9	US-10-174-572-296	Sequence 296, App
25	111	4.6	323	9	US-10-174-579-296	Sequence 296, App
26	111	4.6	323	9	US-10-174-582-296	Sequence 296, App
27	111	4.6	323	9	US-10-174-588-296	Sequence 296, App
28	111	4.6	323	9	US-10-175-739-296	Sequence 296, App
29	111	4.6	323	9	US-10-175-740-296	Sequence 296, App
30	111	4.6	323	9	US-10-175-743-296	Sequence 296, App
31	111	4.6	323	9	US-10-176-488-296	Sequence 296, App
32	111	4.6	323	9	US-10-176-492-296	Sequence 296, App
33	111	4.6	323	9	US-10-176-747-296	Sequence 296, App
34	111	4.6	323	9	US-10-176-750-296	Sequence 296, App
35	111	4.6	323	9	US-10-176-985-296	Sequence 296, App
36	111	4.6	323	9	US-10-176-987-296	Sequence 296, App
37	111	4.6	323	9	US-10-176-991-296	Sequence 296, App
38	111	4.6	323	9	US-10-176-992-296	Sequence 296, App
39	111	4.6	323	9	US-10-176-993-296	Sequence 296, App
40	111	4.6	323	9	US-10-184-658-296	Sequence 296, App
41	111	4.6	323	9	US-10-173-695-296	Sequence 296, App
42	111	4.6	323	9	US-10-173-697-296	Sequence 296, App
43	111	4.6	323	9	US-10-173-705-296	Sequence 296, App
44	111	4.6	323	9	US-10-174-576-296	Sequence 296, App
45	111	4.6	323	9	US-10-174-585-296	Sequence 296, App

ALIGNMENTS

RESULT 1
US-09-924-358-22
Sequence 22, Application US/09924358
Patent No. US20020107376A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
TITLE OF INVENTION: 58764,
FILE REFERENCE: 38155-20034.00
CURRENT FILING DATE: 2001-08-06
CURRENT APPLICATION NUMBER: US/09/924,358
PRIOR APPLICATION NUMBER: US 60/229,300
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatsSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 191
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid
US-09-924-358-22

Query Match 7.6%, Score 181, DB 10, Length 191;
Best Local Similarity 32.7%, Pred. No. 1e-08, Indels 24, Gaps 9;
Matches 65; Conservative 32; Mismatches 78;
259 REHRESIGVSEEDLLFAIINSVRSGODLFLQAFYQALQLOHEKLVPIIHAVVGS 318
2 REEIRKTKGIKEDKTIILFVGRVLPEKGIIDLIEAFKLLKKKPKLTKLN-PULKLYIVGG 60
319 DVNAQT-KPEITLRFVFNKTHDRVHFNKTLAVP-----YLAADIVVONSQGRG 370
61 PYDESDDEEDLKLAKLGLLENVIF----LGVPEDELPELYKSA-DVVLPS--RY 113
371 ECFGRITLMAKPLVLTG-MAAGTTEIVLDGSTGLLHPAGKEGVAPLAKIYVTLASHA 429
114 EGFIVLEMAKGLPVATNCVGGIPEVVKDGETGLLVEPGQDPA-LAEAIETKLDE 172


```
Db 115 VRIIVYAHGRSTETVLGKKKILVFLRYQL-FAHQTLAVSHLKKKLEDAHPFKR 173
Qy 227 LKIQMPQYVHLG---NSKELMEVADNARRVLEHNI--RESLGRSDDLFAIINS 280
Db 174 VCV-----LCYGSMMNSVEL-----DRPRVSLKEKISARNALNIPSSKYVGVGR 219
Qy 281 VSRGKGODFLQAFYQALQLOIHEKLVKPRIHAVVGSVDNAQTKFETQLRDPVVK--- 336
Db 220 INKDGGLL-----AALTKEAFTRLRHLHLIIE-----LEDDLREAFIKLVNE 266
Qy 337 --NTIHDVHFVNTKLAIVAPYLAIDVLYONSQGRGECFRITTEMAFKLPVLTAGG 394
Db 267 GQVLTIGWIDPPEE-----LAAVDVLIHPLOREG--LQMSLLEQVMGVPLTAVTG 318
Qy 395 TTEIVLDGSLGLHPHAGKEGAPLAKNIVRLASHAEORVSGEKGYGVKEMFENHMA 454
Db 319 TVDAVTSSEGGFF--ADDDSVESWVSKIDLVDSPKLRDMGRAGROFVSARFVRDVA 376
Qy 455 RIA 457
Db 377 RFS 379

RESULT 5
US-09-924-358-5
; Sequence 5, Application US/09924358
; Patent No. US20020107376A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Macbeth, Kyle
; TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
; TITLE OF INVENTION: 58764,
; TITLE OF INVENTION: NOVEL HUMAN TRANSFRASE FAMILY MEMBERS AND USES THEREFOR
; FILE REFERENCE: 38155-2003.4.00
; CURRENT APPLICATION NUMBER: US/09/924,358
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/229,300
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-924-358-5

Query Match 5.6%; Score 134.5; DB 10; Length 416;
Best Local Similarity 21.7%; Pred. No. 0.00062;
Matches 92; Conservative 61; Mismatches 193; Indels 77; Gaps 17;

Qy 79 VILVSHSLSGSPILLMEALFRLRHGQV-VWITNQRSGENTDVTYSLERMLNHQVQ 137
Db 17 VLFVHPDLGVGHERVLDLALALQARGCSVKIWTNH-----YDGHGCPAES--R 64
Qy 138 VLPR-----QGEAVDIALKADVLILNTAV--AGKMLDEVLKDHVPVLEP- 180
Db 65 ELPRCAGDWLPRGLGMDGGRGAACAVVRVFLALVYVFLADEDFEDVVDQVSAQIPVF 124
Qy 181 -----KILMWIH-----EMRGHYKPYEVYVHLPFVAGAMIDSHTTAEVNNSRSTRL 227
Db 125 RLARRRKKILPYCHPDDLTKRSPFLAKIYAPIDW-----IEYVTTG-----MADCI 173
Qy 228 KIQMPQYVHLGNSKEIMEVAED-----NVAR--RYLREHIRESLGVASEDLFPALIN 279
Db 174 LVNSQFTAAVFKETFLSLSHIDPVLVPSLNTVTSFDSVVPKUL-DOLVPRGKKFLLLSIN 232
Qy 280 SVSRGKGODFLQAFYQALQLOIHEKLVKPRIHAVVGS---SDVNAQTKFETQLRDPVVK 336
Db 233 RYERKKULTLALALALQJLRGLTSGDWE--RYHLIYAGGYDBRLVNEVHYQELKKVYQO 290
Qy 337 NTIHDVHFV---NKTILAVPIYLAIDVLYONSQGRGECFRITTEMAFKLPVLTAGG 393
```

```
Db 291 SLLGQVTVFRRSPDKOKISLHSCCTCLVTPS---NHFQIVLEEMVQCEVIAVNSG 347
Qy 394 GTTEIVLDGSLGLHPHAGKEGAPLAKNIVRLASHAEORVSGEKGYGVKEMFENHMA 453
Db 348 GPLESIDHSTVGLCPDPVHFSEAIKFIIRPS---LKAHMLAGARVKEKFSSEAF 404
Qy 454 ERI 456
Db 405 EQL 407

RESULT 6
US-09-738-626-4738
; Sequence 6, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4738
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-4738

Query Match 5.6%; Score 133.5; DB 9; Length 409;
Best Local Similarity 23.7%; Pred. No. 0.00075;
Matches 86; Conservative 52; Mismatches 160; Indels 65; Gaps 13;

Qy 133 NHGVOVLPARGQEAVIDIALKADVLILNTAVAGKMLDEVKDH-VKRVL-----PKILMW 185
Db 81 NPAIKTL-STGLRMAEKAANNVDVSHHTWYAGLGHILAAELHGIPIHVAHTASLEPDRPK 139
Qy 186 IHMRGHY-----FKVEYVYVHLPFVAGAMIDSHTTAEVNNSRSTRLKIQMPQYV 237
Db 140 REQLGGGYVSSWGSKAMAYADAVIAVSAHMDSIILAA-----YPIEDPNRVV 190
Qy 238 HLGNSKEL-----MEVAEDNARRVLEHNIRESLGRSDDLFAIINSVSRGKGODFL 291
Db 191 LNGIDTLEWQRPPTFDLAEDESVLR-----SLGVDPRPPIVAFVGRITRQGVGHLI 241
Qy 292 QA-----FYQALQLOIHEKLVKPRIHAVVGSVDNAQTKF--TQLRDPVYNTIHDVHF 345
Db 242 KAAALFDESQVQVLCAGAPDTPFIARFTTALVEELQAKRGIFWQVDMIGDKIOE----- 297
Qy 346 VNKTILAVPIYLAIDVLYONSQGRGECFRITTEMAFKLPVLTAGGTEIVLDGSLG 405
Db 298 -----ILTAADTFVCS--IYEDLIYVLEMMACNTAVASVSGIPEVVDGDTG 346
Qy 406 LHPHAGKEGAPLAKNIV---RLASHAEORVSGEKGYGVKEMFENHMAERIAAVLK 461
Db 347 ALVHYDENDVTFERDIALAENVKVADRETAKTGLAGREERAINDSWATTIAOQTIDVVK 406
```

Qy 462 DVL 464
Db 407 SLM 409

RESULT 7

US-09-738-626-3951
Sequence 3951, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENDO, AKIHIRO
APPLICANT: IKEDA, MASATO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 3951
LENGTH: 418
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3951

Query Match 5.4%; Score 128.5; DB 9; Length 418;
Best Local Similarity 21.1%; Pred. No. 0.0022;
Matches 82; Conservative 47; Mismatches 151; Indels 109; Gaps 16;
Qy 64 AAGSSPLGMRKLV--LVSHSLSGPGLLMELFLR-----HVSQVWIT 113
Db 86 AAFGTGMLSTFRREKTYDLHSHYWSG-----QVGLRLDMRIPLIHTAHLAAVK 139
Qy 114 NORSETNDVYSLLEHRLMHGVLPARGQAVDIALKADLVLTAVAGKVLDPVKD 173
Db 140 NS-YRDSDBPPESEARRICE-----QQLVD--NADVLAVNTQEE----- 175
Qy 174 HVPKVLPRILMWHKRGHYFVEVVKLFPVAGAMIDSHTTAAYMNSRTSDRLKIQMPQ 233
Db 176 -----MODLMHHY-----DADPDRISSVSP- 195
Qy 234 TVVHLGNSKELMEVADNVARVLRHRESLGVSEDLFAIINSVSRKGGODLFLQA 293
Db 196 -----GADVLYSPGNDRATERSRR-----LGRIHTKVVAVGLQFPKGPVLKKA 244
Qy 294 FYQALQILQHEKTKVRIHAVVSGDVNAQTKETQLDPPVAKTTHDRVHPN---KTL 350
Db 245 VAALFDDPPRNLRV-----IICGSPGNATPDT-YHMAEELGVKRIKRLDRPPESE 298
Qy 351 AVAPYLAIDVLVQNSQGRGCFRITTEAMAFKLPVLTAAAGTTEIVLDSTGLLHPA 410
Db 299 LVAVYRAADIIVAVPSF---NESFGLVAMEAOAGTFVIAAIVGLPLFAVABEGTGLL--- 352
Qy 411 GKEGVAP--LAKVIYRLASHAEQVSWGE 437
Db 353 -VDGSHAMADALATLDDDETRIRWGE 380

RESULT 8
US-09-738-626-5896

Sequence 5896, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENDO, AKIHIRO
APPLICANT: IKEDA, MASATO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5896
LENGTH: 381
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5896

Query Match 5.3%; Score 127; DB 9; Length 381;
Best Local Similarity 25.2%; Pred. No. 0.0027;
Matches 55; Conservative 33; Mismatches 82; Indels 48; Gaps 7;
Qy 263 RESIGVASEDLFAIINSVSRKGGODLFLQAFYQALQILQHEKTKPRIHAVVGS----- 318
Db 169 RKKGFPDTPPVIAKNSLIVPRKQDSLIKAMPQVY-----AAPDAQLITVSGRYE 241
Qy 319 -----DVNAQTFETQLRDPVVKNTIHDRVHPNKTALAVPYLAIDVLVQNSQGRG 370
Db 242 STLRRLATVSGNVKFLGRLEYODMINT-----AAADIPAMPARTRG 284
Qy 371 -----ECFGRITTEAMAFKLPVLTAAAGTTEIVLDSTGLLHPAKGE--VAPLAKIV 423
Db 285 GGLDVEGLGIIVLELQAGVPLVAGTSGAPETVTP-ATGLV---VBSGDVXLSELII 339
Qy 424 RLASHAEQVSWGEKGYKRYKEMEHMAERIAVLK 461
Db 340 ELDDPIRRAAMGAGRAHVAEAMSWIEIGERTLNIILQ 377

RESULT 9
US-09-815-028-5
Sequence 5, Application US/09815028
Patent No. US20020068823A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel A.
TITLE OF INVENTION: 33877 AND 47179, NOVEL HUMAN
FILE REFERENCE: 10448-033001
CURRENT APPLICATION NUMBER: US/09/815, 028
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: US 60/191, 964
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 416
TYPE: PRT
ORGANISM: Homo sapiens
US-09-815-028-5

```

QY      319  DVAQOTKFETOLRDFVVKVNTIHDVHFNVKTLAFAVLAIDAIVLVONSQSGECFGRIT 378
Db      236  D--KPKRFA-----LAEKGVRSNVHFPSGRNDVSELMMAADLIHHPAY--QEAAGVITLL 287

QY      319  EANAFLKPYLTAAAGCTTEIVLDGSG--LLHPAGKGVAFPLAQNIVRLA-SHAQFQVSM 435
Db      288  EALTAGLPVLTAAVCGAAHYADANCGTIVLAEFPSSOBQ---NEVLRKALTQSPLEMMAM 343

QY      436  GE--KGVGRVKEMFMEHHMARIAYVL 460
Db      344  AENARHAYADTODLY---SLPEKAADI 367

```

RESULT 12
US-10-174-590-296
; Sequence 296, Application US/10174590
; Publication No. US2003008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

```

? TITLE OF INVENTION: ACIDS ENCODING THE SAME
? FILE REFERENCE: P3430R1C42
? CURRENT APPLICATION NUMBER: US/10/1174,590
? CURRENT FILING DATE: 2002-06-18
? Prior application removed - See File Wrapper or Palm
? NUMBER OF SEQ ID NOS: 612
? SEQ ID NO 296
? LENGTH: 323
? TYPE: PRT
? ORGANISM: Homo Sapien
? US-10-174-590-296

```

Query Match 4.6%; Score 111; DB 9; Length 323;
Best Local Similarity 22.6%; Pred. No. 0.062;
Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11.

```

0Y 181 KILMWH-----EMRGHYFKEVEYKHLPRVAGAMIDSHITTAEYWNRSRTPRLKIMQPOT 234
    |||::| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 39 KILFYCHFPDLLTKRDSFLKRLYRAPIDW-----IEEYTTG-----MADCIIVNSQFT 87

```

27 233IVVLLGNSKELMEVAED-----NVAK--KVLREHIRESLGVRSEDELLFAIINSVSRGKG 286

D5 88 AAVFKETFKSLSHIDPDVLYPSLANTSPDSVPEKL--DVLVPGKGKKLLLSINRIYERKKK 146

147 LTLALBALVOLGRILTSQDWE--RVHLIAGGYDERVLLENVEHYEQLKKMVGQSDLGQYV 204

Db 205 TFLRSFSDKKIKSLHSCTCVLYPPS---NEHGIVPLEAMTMQCPVIAVNSGGPLESID 261

Db 262 H S V T G F L C E P D P V H R S E A I E K F I R P S --- L K A T M G L A G R A V K E K E R S P E A F T E Q L 314

US-10-176-758-296
Sequence 296, Application US/10176758
Publication No. US20030008353A1

```

; APPLICANT: Baker, Kevin P.
;
; APPLICANT: Chen, Jian
;
; APPLICANT: Desnoyers, Luc
;
ADDITIONAL: 000-000-0000

```

/ APPLICANT: Godowski, Paul J.
 / APPLICANT: Gurney, Austin L.
 / APPLICANT: Pan, James
 / APPLICANT: Smith, Victoria

INVENTOR: Macanabe, Colin A.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C104
CURRENT APPLICATION NUMBER: US/10/176,758
CURRENT FILING DATE: 2002-06-21

```

; NUMBER OF SEQ ID NOS: 612
;
; SEQ ID NO 296
;
; LENGTH: 323
;
; TYPE: PPT

```

Query Match	Score	DB	length
US-10-176-758-296	4.6%	9	323
Organism: homo sapien			

Matches	67;	Conservative	46;	Mismatches	143;	Indels	40;	Gaps	11;
0y	181	KILMWIH-----EMRGHYFKVEYKALPFAVGAMIDSHTTAAYNSRFSDRLEKIOWBOT	234						

Dd 39 KILFYCHFPDLLTKRDSFLKRLYPADW-----IEEYTTG-----MADCLIVNSQFT 87

Oy 235 YVHNLNSKSELMENVAEO-----NYAR--RULREHIRESLGVSREDDLLFIINVSNRKG 268
 Db 88 AANPKETFKSLSHIDPVLVPSLANTSDSVBEKI--DVLVPKQKFKLLISTNRYEKKN 146
 Oy 267 QDLFLQAFYQALQIIOHEKLVKPRIHAAVVG--SDVNAQTKFETQDLRFVVKNTIHDRV 343
 Db 147 LTLALBALVQJRGRLTSQDWE--RVHLLIVAGYDERVLENEVHQELKMMVQOQSDLGYY 204
 Oy 344 HFV---NKLNAVAYTLAIDLVLVNSQGBRECGRTITTEBMAFKLPIYLGAAQGTTEVL 400
 Db 205 TFLRSPSDKQKISLHSCVCLVTPRS--NEHGIVLPLEMMYMCPIAVNSGGLESID 261
 Oy 401 DGSFTGLHPAGKEGVAPLAKIVRIASHABQRVMSQKGYGRVKKEMENHHAERI 456
 Db 262 HSYTGLCEPDPIVAFSALKEKIRERS---LKAIVMLAGRAKRVKERSSEALTEEDV 314

RESULT 14
US-10-175-737-296
: Sequence 296, Application IIS/10175737

```

; INFORMATION NO: 00200900130361
;
; GENERAL INFORMATION:
;
; APPLICANT: Baker, Kevin P.
;
; APPLICANT: Chen, Jian

```

; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT:

; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang Zemin

FILE REFERENCE: P3430R1C50
TITLE OF INVENTION: ACIDS ENCODING THE SAME
CLASS OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
CURRENT APPLICATION NUMBER: US/10/175.737

```

; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 296

```

```

; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-296

```

Query match 4.6%; Score 111; DB 9; Length 323;
Best Local Similarity 22.6%; Pred. No. 0.062;
Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11.

Db 39 KILPYCHPDDLTKRDSFLKRLYRAPIDW-----IIEYTTG-----MADCILVNSQFT 87

D_b 88 AAVFKETFKSLSHIDPVLYPSLNVTSPDSVPEKL-DDLVPKGKKLLLSINRYERKKK 146

D0
147 LTLALVALQRLGRITSQDWE--RVHLIVAGYDERLLENVEHYQELKKNVQQSDLGQYV 204

344 HEV---NKTIAVA DVI AATDVH VONSGRGECGCDTTPMAEYK DIT GMAACGGMTIIV 400

[illegible]

Db 262 HSYVIGFLCEPDYPVHFSEAIKFIREFS---LKATMGLAGARVRKESPSPEAFTEQL 314

RESULT 15

US-10-173-706-296

; Sequence 296, Application US/10173706
; Publication No. US20030022293A1

; GENERAL INFORMATION:

APPLICANT: Baker, Ke

APPLICANT: Chen, Jian

APPLICANT: Cneil, Vlad
APPLICANT: Desnoyers

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P3430R1C7

FILE REFERENCE: P3430RLC/
CURRENT APPLICATION NUMBER: US/10/173.706

CURRENT APPLICATION NUMBER: US/10/173,706
CURRENT FILING DATE: 2002-06-17

CURRENT FILING DATE: 200

;
Prior Applic

NUMBER OF SE

SEQ ID NO 296

LENGTH: 323

TYPE: PRT

ORGANISM: *NOCTU* sapient
US-10-173-706-296

Query Match 4.6% Score 111 DB 9: Length 323:

Query Match	Score	DB
4.6%	111	9
Best local similarity	22.6%	
Pred. No.	0.062	

Best Local Similarity 22.6%; Pred. No. 0.062;
Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11;

Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11

PUBOIIWZAMNATU DEVA CAMTDCHTAEVUNISPTSDPI YIOMBOT 234

181 KILWIIH-----EMRGHYFKVEYVKHLPFVAGAMIDSHTTAEYWSNRTSDRLKIQMPQT 234

[illegible]

Db 39 KILFYCHPDL LTKRDSFLKRLYRAPI DW-----IEEYTTG-----MADCILVNSQFT 87

Case	Age	Sex	Duration of illness	Onset	Course	Outcome
1	55	M	10 years	1985	Progressive	Death
2	60	F	15 years	1970	Progressive	Death
3	65	M	20 years	1965	Progressive	Death
4	70	F	25 years	1960	Progressive	Death
5	75	M	30 years	1955	Progressive	Death
6	80	F	35 years	1950	Progressive	Death
7	85	M	40 years	1945	Progressive	Death
8	90	F	45 years	1940	Progressive	Death
9	95	M	50 years	1935	Progressive	Death
10	100	F	55 years	1930	Progressive	Death

235 YVHLGNSKELMEVAED-----NVAR--RVLREHIRESLGVRSEDLFAIINSVSRGKG 286

[illegible]

Db 88 AAVEKETEKSLSHIDPDVLYPSLNTSFDSDVPEKL-DDLVPGKKFLLSINRYERKN 146

Search completed: June 30, 2003, 16:35:04
Job time : 55 secs

0
0
7

7
4
三
.

(
(

E
C
C
Z

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2003, 16:24:52 ; Search time 43 Seconds
(without alignments)

1057.479 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2397
Sequence: 1 MAKTPSPFAVAAGRGQPVH.....ERIAVLKQVLRKQSHSRN 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR 73 : *
1: p1r1 : *
2: p1r2 : *
3: p1r3 : *
4: p1r4 : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1478	61.7	402	2	G96784 hypothetical prote
2	1471	61.4	458	2	A86330 hypothetical prote
3	482.5	20.1	670	2	B96564 hypothetical prote
4	207	8.6	382	2	D82889 glycosyltransferas
5	207	8.6	382	2	A82994 glycosyltransferas
6	197	8.2	416	2	G71096 hypothetical prote
7	187.5	7.8	388	2	E72336 probable hexosyltr
8	181.5	7.6	381	2	F71196 probable hexosyltr
9	172.5	7.2	370	2	B82751 lipopolysaccharide
10	166.5	6.9	377	1	A70004 probable hexosyltr
11	166	6.9	382	2	A82368 hypothetical prote
12	161	6.7	405	2	B96817 hypothetical prote
13	156	6.5	416	2	B95920 probable membrane-
14	154.5	6.4	351	1	H64446 probable hexosyltr
15	153	6.4	381	2	C41317 probable lipopolys
16	151.5	6.3	411	2	H69158 LPS biosynthesis R
17	151	6.3	375	2	C83860 hypothetical prote
18	149.5	6.2	393	2	H72352 lipopolysaccharide
19	149	6.2	381	2	A80973 lipopolysaccharide
20	148.5	6.2	378	2	H83019 probable glycosyl
21	147.5	6.2	351	2	H97621 lipopolysaccharide
22	147.5	6.2	351	2	A87844 LPS glycosyltransf
23	147	6.1	333	2	F84166 LPS glycosyltransf
24	145	6.0	333	2	H84113 hypothetical prote
25	144	6.0	409	2	S76126 hypothetical prote
26	144	6.0	429	2	AC2227 hypothetical prote
27	142	5.9	220	2	A82360 glycosyl transfera
28	142	5.9	349	2	H87379 glycosyl transfera
29	142	5.9	381	2	A82160 glycosyltransferas

30	141.5	5.9	412	2	T35514 probable glycosyl
31	141	5.9	354	2	A83315 lipopolysaccharide
32	140	5.8	380	2	B97275 glycosyltransferas
33	140	5.8	429	2	A82456 glycosyltransferas
34	138.5	5.8	427	2	B95936 probable glycosylt
35	138	5.8	316	2	F70441 capsular polysacch
36	138	5.8	344	2	C81152 LPS biosynthesis p
37	136	5.7	376	2	A82676 conserved hypothet
38	136	5.7	422	2	A82352 hypothetical prote
39	135	5.6	385	2	D97911 hypothetical prote
40	134.5	5.6	390	1	A75059 probable hexosyltr
41	133.5	5.5	230	2	G81870 probable glycosyl
42	132.5	5.5	417	2	A82359 hypothetical prote
43	131	5.5	364	2	H75466 lipopolysaccharide
44	131	5.5	379	2	S77338 LPS glycosyltransf
45	131	5.5	430	2	A12455 glycosyltransferas

ALIGNMENTS

RESULT 1

G96784 Hypothetical protein F1B16.5 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_rev1500 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G96784

R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Matz, R.; Marzalli,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96784

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-402 <STO>

A:Cross-references: GB:A8005173; NID:gi1020445; PIDN:AA013070.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1B16.5

A:Map position: 1

Query Match 61.7% Score 1478; DB 2; Length 402;

Best Local Similarity 74.5% Pred. No. 9.6e-102;

Matches 283; Conservative 53; Mismatches 44; Indels 0; Gaps 0;

QY	89	SGGPIILMEIAFLRHVGSQVWITNORSQETNDYTSLEHMLAHGVQVLPARGDEND	148
DB	18	SRGPIILMEIAFLRHVGSQVWITNORSQETNDYTSLEHMLAHGVQVLPARGDEND	77
QY	149	IALKADLVITAVAGKMLDPVKDHPKVLPKILMIMHEMRGHFKVEYKHLPPVAGA	208
DB	78	TSALKADLVITAVAGKMLDPVKDHPKVLPKILMIMHEMRGHFKVEYKHLPPVAGA	137
QY	209	MIDSHTAAVWNSRTSRLKIQMPQTYVHAGSKELMEVAENNAKRYLEHRESLSQ	268
DB	138	MIDSHTAAVWNSRTSRLKIQMPQTYVHAGSKELMEVAENNAKRYLEHRESLSQ	197
QY	269	RSEDLFAITNSVSGKODLPLOAFYQALQIOHEKIKAPRIHAVVGSQVNAQTFFET	328
DB	198	RSEDLFAITNSVSGKODLPLOAFYQALQIOHEKIKAPRIHAVVGSQVNAQTFFET	257
QY	329	QLRDPVVKNTIDHRYFNKTLAVAPYLAIDVLYVNSQSGRCERTITAMAFLKPLV	388
DB	258	ELRPFREKKLENFVFNKTLAVAPYLAIDVLYVNSQSGRCERTITAMAFLKPLV	317
QY	389	GTAAGGTIVDGSGLHPAKESGVAFLKNTIVRLASHAEORVSMGKYGKEMFM	448

Db 318 GTAAGTMEIVNGTGLTSHSAGKEGVIPLAKNIVKLATQVELRLMGKNGYERVKEMFL 377
Oy 449 EHHMARIATVAKDVLKRSQ 468
Db 378 EHHMARIATVAKDVLKRSQ 397

RESULT 2

A66330
hypothetical protein F69.24 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 31-Dec-2001
C:Accession: A66330
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: A66330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-458 <STO>
A:Cross-references: GB:AE005172; NID:g10086496; PID:AMG12556.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 61.4%; Score 1471; DB 2; Length 458;
Best Local Similarity 62.8%; Pred. No. 3.8e-101;
Matches 304; Conservative 65; Mismatches 73; Indels 42; Gaps 6;
Oy 1 MAKTPSPAAVAVAGRGVPHNRFTLLLVAAASATAGFLRGALRDPDRCGR--- 56
Db 1 MAKTPSPAAVAVAGRGVPHNRFTLLLVAAASATAGFLRGALRDPDRCGR--- 56
Oy 57 -----DPAALNTAVASGS--PLGPMRSKVLVNHSELSISGGPILMLATLRLHVGSOVY 110
Db 51 REKEDNSDIKIOSVSSLNPLLEPMKSKVLVLSHSELSISGGPILMLATLRLHVGSEVY 110
Oy 111 WITNORSQETNDVYSLERHMLNFGVYLPARGQEAVIDALKADLVLTNTAVAGKMLDPV 170
Db 111 WITNORSQETNDVYSLERHMLNFGVYLPARGQEAVIDALKADLVLTNTAVAGKMLDPV 170
Oy 171 LKDHVPVLPKILMWHMIRGKHYFVYKGLPFAVAGMIDSHSTAAVYNNSTSDRLKIQ 230
Db 171 LKDHVPVLPKILMWHMIRGKHYFVYKGLPFAVAGMIDSHSTAAVYNNSTSDRLKIQ 230
Oy 230 MPTQYVYVHLGNSKELMEVAEDNVARVRLREHRESLGVASBDDLFAIINSVSRGKQDLF 290
Db 230 MPTQYVYVHLGNSKELMEVAEDNVARVRLREHRESLGVASBDDLFAIINSVSRGKQDLF 290
Oy 291 LQAFYQALQIQH-EKLAKPRTHAVVVGSDVNAQTKEFTQLRDPVYVKNTHIRVFNKT 349
Db 270 LRAHESLAKYIKETKLEVPYMHAVVVGSDMSQTKETELRNVDPMKIQIYHFNKT 329
Oy 350 LAVAPYLAIDVYVNSGREGCEGRTTETANAFLPYLGTAACTTEIVLDGTTGLHP 409
Db 330 MKVAPYLAIDVYVNSGREGCEGRTTETANAFLPYLGTAACTTEIVARRTGLLHN 389
Oy 410 AGEKGVAPLAKNIVLASHAEQVSMGEKGVGKGVMEHMAHIAVILKDVLRKSOE 469
Db 390 TKGQGVLPPLAKNIVKLATNIVMQRNTMGKGYERVKEMFLRHHSRHSIATVLEVLQAKI 449
Oy 470 HSR 473
Db 450 HSR 453

RESULT 3

E96564
hypothetical protein F6D8.36 (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 31-Mar-2001
C:Accession: E96564
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: E96564
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-670 <STO>
A:Cross-references: GB:AE005173; NID:g5903062; PID:AMD5621.1; GSPDB:GN00141
C:Genetics:
A:Gene: F6D8.36
A:Map position: 1

Query Match 20.1%; Score 482.5; DB 2; Length 670;
Best Local Similarity 29.2%; Pred. No. 9.7e-28;
Matches 148; Conservative 77; Mismatches 173; Indels 109; Gaps 13;
Oy 38 STAGFLIR--GALRD-----PQGRGDPALNTAVASGSPGFMK---SKVL 80
Db 193 ATYKLGIFPGSLEIKVLEWSPHRSRGTCRKS-----FKLIWSSRFV 237
Oy 81 LVSHSELISGCPPLMLATLRLHVGSOVYVWITNORSQETNDVYSLERHMLNFGVYLP 140
Db 228 LRFELSLMTGALISMELASLSCATVSAVLSRG-----GLMELSRRIKIVE 290
Oy 141 ARQGEAVDIALKADLVILNTAVAGKMLDPVLPKILMWHMIRGKHYK----- 195
Db 291 DKGLSLFKYAMKADILLAGSVCTSWIDQYMH-PAGGSQIAMVMEKREYDPRAKV 349
Oy 196 VEYVKHLPVAGAMIDSHSTAAVYNNSTSDRLKIQMPTQYVYVHLGNSKELMEVA----- 249
Db 350 LDRKMLIFL-----SESQRMLTWCEEHKILR--SQVYIPLSVDELAPYAGIPSS 402
Oy 250 -----ENVARVLRHRESLGVASBDDLFAIINSVSRGKQDLFQAFYQALQ- 299
Db 403 LNTPTLSPEKMKVRKQILRESVTELTGIDSDMLWSLSINFTKGQILLLESIALSL 462
Oy 300 -----LQHEKLVKPRTHAV-----VYGS 318
Db 463 RQGESQRNHKGIIRKAVSLSKHRLRGSSRQMSVSLTLDNGLRKQELKVLGSSVS 522
Oy 319 DYNAQKKEFTQAPDFVYVKNTHIRVFNKTIAVAPYLAIDVYVNSGREGCEGRTT 378
Db 523 KSKNKGIVKEMLSLNSGMLSKSVMTPTATTRVASLYSAADVYVNSGREGCEGRTT 582
Oy 379 EANAFLPYLGTAACTTEIVLDGTTGLHPAGKEGVAPLAKNIVLASHAEQVSMGEK 438
Db 583 EANAFLPYLGTAACTTEIVLDGTTGLHPAGKEGVAPLAKNIVLASHAEQVSMGEK 438
Oy 439 GYGRVYKEMMEHMAHIAVILKDVLR 465
Db 643 GRGVKEMMEHMAHIAVILKDVLR 669

RESULT 4

D98289
hypothetical protein AGR.L2541 (imported) - Agrobacterium tumefaciens (strain C58, Cerec
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #ext_change 11-Jan-2002
C:Accession: D98289

R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Molism, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappae, C.; Markelz, B.; Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*
A:Reference number: A97359; PMID:11743194
A:Accession: D98289
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <KIR>
A:Cross-references: GB:AE007870; PIDN:AAK89838.1; PID:g15159774; GSPDB:GN001870
C:Genetics:
A:Gene: AGR_L_2541
A:Map position: linear chromosome

Query Match 8.6%; Score 207; DB 2; Length 382;
Best Local Similarity 27.5%; Pred. No. 1.1e-07;
Matches 68; Conservative 46; Mismatches 109; Indels 24; Gaps 7;

QY 220 NSRTSDRLKI-----QMPQTVVHLGNSKELMEVADVARVLRHIRESLGVASEDLIF 275
DB 154 NSEETGAFTEGGEADKRVIVNGFDPKAKLHDQMAAR-----LRAELGLQPOPLV- 207
QY 276 AINSVSRGKGQDLFLQAFYQALQIHEKLVPRIHAVVVGSDVAQTKETQLADPVL 335
DB 208 GLFRLSEWKGQHVFLDL-AMEGVQ-----AVIVGALFGQBAVEARIREQAS 256
QY 336 KTIHHRVHFVVKTLAAVPLAIDVLVNSGSGEGCGRTTEMAFKLPVLGTAAAGT 395
DB 257 RLGLDGRVETGFRSDPBEIAMSMDVAHTSI-VAEPGRVIVVEAMCGRPVATRGGSV 315
QY 396 TEIVLDSTGLHPACKEGVAPLAKNIVELASHAEORVSMGEGYGRVKEFMENHMER 455
DB 316 TEIIRDETGLVLPQG--DASALAAALGTLISDPALAQRLQSGREDVSDRFLOETCRS 373
QY 456 IAAVLKD 462
DB 374 VSALLTE 380

RESULT 5
AF2994
glycosyltransferase Atu3560 (imported) - *Agrobacterium tumefaciens* (strain C58, Dupont)
C:Species: *Agrobacterium tumefaciens*
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF2994
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.;
A>Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2994
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <KIR>
A:Cross-references: GB:AE008689; PIDN:AAI44372.1; PID:g17741967; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3560
A:Map position: linear chromosome

Query Match 8.6%; Score 207; DB 2; Length 382;
Best Local Similarity 27.5%; Pred. No. 1.1e-07;
Matches 68; Conservative 46; Mismatches 109; Indels 24; Gaps 7;

QY 220 NSRTSDRLKI-----QMPQTVVHLGNSKELMEVADVARVLRHIRESLGVASEDLIF 275
DB 154 NSEETGAFTEGGEADKRVIVNGFDPKAKLHDQMAAR-----LRAELGLQPOPLV- 207
QY 276 AINSVSRGKGQDLFLQAFYQALQIHEKLVPRIHAVVVGSDVAQTKETQLADPVL 335

DB 208 GLFRLSEWKGQHVFLDL-AMEGVQ-----AVIVGALFGQBAVEARIREQAS 256
QY 336 KTIHHRVHFVVKTLAAVPLAIDVLVNSGSGEGCGRTTEMAFKLPVLGTAAAGT 395
DB 257 RLGLDGRVETGFRSDPBEIAMSMDVAHTSI-VAEPGRVIVVEAMCGRPVATRGGSV 315
QY 396 TEIVLDSTGLHPACKEGVAPLAKNIVELASHAEORVSMGEGYGRVKEFMENHMER 455
DB 316 TEIIRDETGLVLPQG--DASALAAALGTLISDPALAQRLQSGREDVSDRFLOETCRS 373
QY 456 IAAVLKD 462
DB 374 VSALLTE 380

RESULT 6
G71096
hypochemical protein PH1035 - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: G71096
R:Kawabuchi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon
A:Reference number: A71000; MIMD:98344137; PMID:9679194
A:Accession: G71096
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-416 <KAM>
A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30133.1; PID:d1031076; PID:g325745
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by Genbank
C:Genetics:
A:Gene: PH1035

Query Match 8.2%; Score 197; DB 2; Length 416;
Best Local Similarity 25.5%; Pred. No. 6.7e-07;
Matches 104; Conservative 56; Mismatches 174; Indels 74; Gaps 18;

QY 87 SLGGFPL-LMELAFILRHVGSQVVTINQSRQETNDVYSLHRLMLNHGVQLVPARGOE 145
DB 50 SPFGGVAELHSLVPLRLSIGIEARVPTEGTFEFVTKTHNL-----QGNB 99
QY 146 AVDIALKADVLIN-TAVACKMLDPVLKDHVPKVLKILMWHENRGHYFVETVYGLPF 204
DB 100 SLCTPEKKELYLVNENSKFIDSSFDV-----LVHDPALALIEFEKKSFW 150
QY 205 VAGAMID-SHTTAAYN-----SRSDRLKIQMPQTVVHLGNSKEL-----EVAEDN 252
DB 151 LMRCHIDLSNRRFEFLRRFVEXDYIFHLPEYVQPELDRNAVIMPPSIDPLSEKN 210
QY 253 VARR--VLRHIRESLGVASEDLFLAIIINSVR--GKGQDLFLQAFYQALQIHEKLV 306
DB 211 VELKQTEILR--ILERPVDPEK--PIITQVSRFPWKQ-----IFDVIEIRKVE 258
QY 307 KVPRIHAVVG--SVNNAQTKETQLR--DPVK--NTIHRVHFVVKTLAAVPL 355
DB 259 KIVGQVLILGVMAHDPBEGMIVFEKTLRKIGEDYDKVLTNLIGVHARVN-----AF 312
QY 356 IAAIDVLVNSGSGEGCGRTTEMAFKLPVLGTAAAGTTEIVLDGSGTGLHPACKEGV 415
DB 313 QRSADVILQMSIRSG--FGLTVEAMWKKQKPVIGRAVGGIKQIVDGEFTLVADANEAV 370
QY 416 APLAKNIVELASHAEORVSMGEGYGRVKEFMENHMERIAAVLVKD 463
DB 371 ----EKVLYLKHPEVSKEMGAVAKERVKNPITTKMERVYLDLINSL 414

RESULT 7
E72354
Probable hexosyltransferase (EC 2.4.1.-) TM0622 - *Thermotoga maritima* (strain MSB8)
C:Species: *Thermotoga maritima*

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C/Accession: E72354
R/Name: K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.,
C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A/Reference number: A72200; MUID:99287316; PMID:10360571
A/Accession: E72354
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1388 <ARN>
A/Cross-References: GB:AE001736; GB:AE000512; NID:g4981138; PIDN:AD035706.1; PID:g498114
A/Experimental source: strain MSB8
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 7.8%; Score 187.5; DB 2; Length 388;
Best Local Similarity 23.5%; Pred. No. 3e-06;
Matches 89; Conservative 58; Mismatches 136; Indels 93; Gaps 13;

136 VQVLPARGQAEVADTALAKADLVILNTAVAGKWLDPVLKDHVPKVLPRKIIMTWHEMRGHYK 195
Db 47 VEKLTSGKYQVYTVLDYE-----ALNP-----KYIRRLRAIKMKRYTL 89
Qy 196 VEVYK-----HLPFVAGAMDS-----HTYE-----YNSRTSRLKI 229
Db 90 LBEIRPDITSHLSALMTALITPLLCRIPKVTHTHVAEKAKGITRFPN-----RI 142
Qy 230 QMPQTYVYVHLGNSKEIMEVADNVARRVLRHIRESLGV-----RSEDLFPKI 277
Db 143 AKFPKFPVPIVSIQSEVASESVKLYGRKISTEPIVYNGIDQKFSIDQPRKVDROKTLINIV 202
Qy 278 INSVRGQDLPFLQAFYQALQTLQHEKLYKPRHVAWVGSDVNAQTFFEQLRDPVYKN 337
Db 203 AR-LSREKNAHLVRAFSKAVQ-----SCPNIEMLVGDG-----ELRRDIEELVKOL 249
Qy 338 THHDVHFUNKTLAVAPYLAIDVLVONSQGRGECFGRITTEINAAFKLPVLGTAAGTTE 397
Db 250 GLEEVKVFPGVRSIVPELLSQADIFVLSDPYEG--FGVLVAENMAAGLPVIAITAGIGPE 307
Qy 398 IYLDGSGTLLHPAGKGVAPLAKNIYVRLASHAQRSMKMKGYKYMMEHMAERIA 457
Db 308 ILEGGRAGILVP--PKVDALAKAIVELARDEKKAELSDYGRKLVARF----- 355
Qy 458 AVLKQVLRKSGQHSR 472
Db 356 ----DIRRTVREYK 366

RESULT 8
F71196
C/Probable hexosyltransferase (EC 2.4.1.-) PH1844 - Pyrococcus horikoshii
C/Species: Pyrococcus horikoshii
C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C/Accession: F71196
R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hakiawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kuchida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A/Reference number: A71000; MUID:98344137; PMID:9679194
A/Accession: F71196
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-381 <KAW>
A/Cross-References: GB:AP000007; NID:g3236134; PIDN:BA030965.1; PID:g3258282
A/Experimental source: strain OT3
A/Note: this accession replaces an interim accession for a sequence replaced by GenBank
C/Genetics:
A:Gene: PH1844
A:Superfamily: probable hexosyltransferase ytkN

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 7.6%; Score 181.5; DB 2; Length 361;
Best Local Similarity 23.1%; Pred. No. 8.2e-06;
Matches 93; Conservative 67; Mismatches 161; Indels 81; Gaps 15;

OY 98 IAPLRHVSQVAVITNORSQETNDVYSLSHRLNMGVLPARG--QEAVIDALKADL 155
DB 27 IAIKLRRRGHEVGIYTNRPPT-----GKEELKRYGIEILIKIPGISPFPLDNLTVGL 79
OY 156 VLTMTAVAGKMDLPVLCQHVPEKVLPKLWMIHENMGYFKVEYKHLR-----FVAGMID 211
DB 80 -----KSSEINLEPKD-----FDIHS--HHAFTPLSKALKAGMKKEKTLIT 122
OY 212 SHTTAEVWNSRSTSDRLKIQMP--QTYVTH---LQNSKELMEVAEDNVARVL----- 258
DB 123 THSISPAHESKMDLTGFTIPLPKFSYKTSYSHRIIVASFAAKSPTEHTSPVPLIVPQVD 182
OY 259 -----REHIRESLGVRSDDLAFIINVSVRGQDPLQAFYQALQIHEKLRVP 309
DB 183 DERFPAPARDEKIKAKFGLEGVNVLY--VSRMSYKRGPHVLNMF-----SITE 229
OY 310 RIHAVVVGSD-----VNAQTEETQLRDVPVQNTIHDHNVHVKTLAVALPYLAIDVYQ 364
DB 230 DATLVWNGSEMLPFLKQITVF-----LGINKKVVENGVPYDDILPEVFRNADVPFLPS 283
OY 365 NSQGRGECFRITTEAPMAFKLPVLGTAGGTTETVLDGSTGLLHPAGKEGVAPLKNIVR 424
DB 284 ISS--SSAFPIYILEPMAASGVPIIATDVGIEPVYKENSAGLIVPQGN--LKLREAIK 338
OY 425 LASHAEQVSMGCKGVGRVCKEMPEHMMERLAALKLVYLRK 466
DB 339 LKKEELRKQYNGNRBSREVESYKMKIVYKIERIYNVLDG 380

RESULT 9
EB82751
1lipopolysaccharide biosynthesis protein Xf0879 [imported] - Xylella fastidiosa (strain 9608)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: EB82751
R:anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MIMD:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: EB82751
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <SIM>
A:Cross-references: GB:AE003927; GB:AE003849; NID:G9105783; PIDN:AAFP3689.1; GSPDB:GN001.1
A:Experimental source: strain 9608
R:Simson, A.J.G.; Reznick, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al-
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H
de-Nero, E.; Docena, C.; El-Dorry, H.; Fachinani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
U:Authors: Ferreira, V.C.A.; Perro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
U.D.; Unqueira, M.L.; Kemper, E.L.; Kitzajma, J.P.; Krieger, J.E.; Kutzmae, E.E.; Laigre
Chado, M.A.; Madalari, A.M.B.N.; Madalari, H.M.F.; Martin, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsubara, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiti, D.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tanaka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.; Za
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: Xf0879

Query Match 7.2%; Score 172.5; DB 2; Length 370;
Best Local Similarity 25.0%; Pred. No. 3.7e-05;
Matches 79; Conservative 48; Mismatches 136; Indels 53; Gaps 12;

Db 113 FAYDKISOAE-----GEICIDISVAQIVTWSVDYIGOTTISREPPSARSKITIVASGVDLKTY 167

QY 247 EAVEDVARVLRHIRESLIGRSEDLFLAIINSVSRGKGDLLFQAFYALOLIQHEKL 306

UD 306 INNOVATIONEN FÜR DIE ZUKUNFT
QY 436 GEK 438

RESULT 12
B96817
hypothetical protein F9K20.16 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 11-Mar-2001
 C/Accession: B96817
 R/Theologian, A.; Ecker, J.R.; Palm, C.U.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso, Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 815-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, U.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani, R.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: B96817
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-405 <STO>
 A:Cross-references: GB:AE005173; NID:g3834314; PIDN:AAC83030.1; GSPDB:GN00141
 C/Genetics:
 A:Gene: F9K20.16
 A:Map position: 1

Query Match 6.7%; Score 161; DB 2; Length 405;
 Best Local Similarity 22.3%; Pred. No. 0.0003;
 Matches 97; Conservative 74; Mismatches 186; Indels 78; Gaps 17;
 Oy 76 SKLVLVSH-ELSLGGPULLMELAFLLRHVGSQV-WITNQRSGEINDVYSLERHMLN 133
 Db 7 SKMNAIITHPDIGIGGAEIRLYDAVAVELASHGHKVIHPTSHHDSRCREETLSGIQYIV 66
 Oy 134 HGCVQLPAR--GQEAVDIALKADVILNTRVAGKGLDPLVDKHYPRKLP-----KI 182
 Db 67 YG-STLPRHIFYRLHAYCAVLCFLVCLILGSSFDVLDVQSVVVLKLKSSKY 125
 Oy 183 LWMH-----EMRGHYFK-VKYYKHLPPVAGAMI--DSHTTAETWNSRTSDRLK 228
 Db 126 VVYCHFPDLLAKHTTLTRRMRYRKEIDFIEQQTGMADILVNSFTASTF-ANPEKLN 184
 Oy 229 IIMPQTVVHLGNSKELMEVADNVARVREHIRESGVSEDLALAINNSRKQD 288
 Db 185 AGGSPRAVLPA-----VNIDFIEPHYKLANLSTNRPERKIND 225
 Oy 289 LFLQAFYQALQIQLHEKLVPRHAAVVG-----SDVNAQTKEFTQLRDFVKNTHDRV 343
 Db 226 LAVSAF-----AIIICKHKOMLSDVTLVAGCGYDEIKENVEYELRLSLAEKGVSDRV 281
 Oy 344 HFV-----NKTLLVAPYLAIDVIVNSQGRGCEGRITTEMAFKLPVIGTRAGT 395
 Db 282 NFITSCGTARNELLS-----SCLCVLYTPTD--EHFQIVPLBMAAYEPVLAICNSGSP 333
 Oy 396 TEIVLDGSTG-LIHPAGKEGVAFLAKNIVLASHAEQVSMGKGVGRVYKEMFEMHMAE 454
 Db 334 VETVNGVNTVYLCPEPBDFFSSAMA-----RLENPFLANRAGARHVVESPSVKTGQ 389
 Oy 455 RIAAVLKDVLRSQE 469
 Db 390 KLNQYLVDVVSSPKE 404

RESULT 13

H93920
 Probable membrane-anchored glycosyltransferase protein Smb21053 [imported] - *Sinorhizobium meliloti*
 C/Species: *Sinorhizobium meliloti*
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: H93920
 R/Finn, T.M.; Weidner, S.; Wong, K.; Buhrmester, U.; Chain, P.; Vorholter, F.J.; Hernan Proo, Natl. Acad. Sci. U.S.A. 96, 9885-9894, 2001
 A>Title: The complete sequence of the 1.683-kb psymb megaplasmid from the N2-fixing endc
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: H93920
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-416 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49032.1; PID:g15140517; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid psymb
 R/Galbert, F.; Finn, T.M.; Long, S.R.; Punler, A.; Abol, P.; Ampe, F.; Barloy-Hubler, P.; Chlu, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federpiet, N.A.; Fisher, R.F.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leblanc, P.; Vanderbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A>Title: The complete genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C/Genetics:
 A:Gene: Smb21053
 A:Genome: plasmid

Query Match 6.5%; Score 156; DB 2; Length 416;
 Best Local Similarity 25.4%; Pred. No. 0.00072;
 Matches 59; Conservative 41; Mismatches 98; Indels 34; Gaps 7;
 Oy 261 HIRSLGVRSNED-----LFLAINSVSRKSGODLFLQAFYQALQIQLHEKLV----- 308
 Db 192 HFRPSEAGASGARTRGVILF--VGRISPEKGLHTLVAFSEVALRFPVDELRIAGPYSP 249
 Oy 309 -----PRTHAVVGSQVNAQTKEFTQLRDPVKNTHDRVH-----VNKTLLAVA 353
 Db 250 LPVDFLTSBSPRVLDKRFYDQMRCKYQOHDELMDHRLRLHRIKLVGNSHKEIVA 309
 Oy 354 PYLAIDIVVNSQGRGCEGRITTEMAFKLPVIGTRAGTTEIVLDGSTGLHPAGKE 413
 Db 310 AYHDA-DIVVNS--LSSEFGISVSGMACGIPVGTGVRGCMSESILDSHTMTLEADAP 366
 Oy 414 GVABLANIVRLASHAEQVSMGKGVGRVYKEMFEMHMAERIAVLAVKDYL 465
 Db 367 G--ELSGALITVLDPPARAGMGTEGRERAVALVSWEARERLRSYVERVSR 416

RESULT 14

H64446
 Probable hexosyltransferase (EC 2.4.1.-) M01178 [similarity] - *Methanococcus jannaschii*
 C/Species: *Methanococcus jannaschii*
 C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
 C/Accession: H64446
 R/Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurelt, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 A>Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: H64446
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-351 <RUT>
 A:Cross-references: GB:U67559; GB:L77117; NID:g1591798; PIDN:AAB99181.1; PID:g1591805; T1
 C/Genetics:
 A:Map position: REV1117459-1116404
 A:Start codon: GTG
 C:Superfamily: probable hexosyltransferase ycxN
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 6.4%; Score 154.5; DB 1; Length 351;
 Best Local Similarity 27.2%; Pred. No. 0.00073;
 Matches 55; Conservative 34; Mismatches 88; Indels 25; Gaps 7;
 Oy 242 SKELMEVADNVAR--VLEHIRESLGVRSNEDLFLAI-INSVSGKQDPLQAFYQAL 298
 Db 146 SKYIKQDLDEMILKRAIVIVNGEILVYNGDVNFGLPFGAFVQKQVLDLIDA----- 200
 Oy 299 QIQHKKKLVPRHAAVVGSDVNAQTKEFTQLRDPVKNTHDRVHFNKTL-AVAPYLA 357
 Db 201 -----IKDDIFNFKLIGG-----KLYKKIENFVKKNLISHIELGRKSPFEVASFMR 248
 Oy 358 AIDVAVNSQGRGCEGRITTEMAFKLPVIGTRAGTTEIVLDGSTGLIHPAGKEGVA 417

DB 249 KCSFLVPS--RSEGFQVAVEGMCSKPVLATRVGGLGEIVDGNGL--AEKNPNPND 304
QY 418 LANNIVRLASHAEQVSMGEK 439
DB 305 LKEKILELLEINNEELRKTLENG 326

RESULT 15

C41317
probable lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56) - *Salmonella* C
C/Species: *Salmonella typhimurium*
C/Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 08-Oct-1999
C/Accession: C41317
R/MacLachlan, P.R.; Kadam, S.K.; Sanderson, K.E.
J. Bacteriol. 173, 7151-7163, 1991
A/Title: Cloning, characterization, and DNA sequence of the rfaK region for lipopolysac
A/Reference number: A41317; MUID:92041612; PMID:1657881
A/Accession: C41317
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-381 <MAC>
A/Cross-references: GB:M73826; NID:g154328; PID:AAA27207.1; PID:g454851
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 6.4%; Score 153; DB 2; Length 381;

Best Local Similarity 23.4%; Pred. No. 0.0011;
Matches 74; Conservative 40; Mismatches 110; Indels 92; Gaps 11;

QY 168 DPLVAKDHPKVLPKILMWHMRGHYFKVEYVKHLPEVAGAMIDSHTTAETWNSRTSDRL 227
DB 133 EPELIPNDAKII-----VPSOFLKAFYEERLPAAAVSIIVNGFCATYKRNPDNL 183
QY 228 KIQMPQTVVHLGNSKEIMEVADNVARRVREHIRESLGVRESDLLFAIINSVSRGKG 287
DB 184 RQO-----LNIADATV-----LLVA--GRISPDGI 208
QY 288 DLFLOAFYALQLIQHEKLVPRILHAVVGVSDVNAQTKEFQLRDFVVKNTIHDRHFVN 347
DB 209 LLLLOAFKQILRTLRNIDL-----VVVG-DPVASRKGE-----KAEYOK 246
QY 348 KTLAVAPVLA-----IDLVONSQSGECFGRITTEMAFKLPVL 388
DB 247 KTLDAKEIGTCIMAGSQSPQOMNFYHIALVIVPSQVE-EAFQVAVEAMAAKAVL 305
QY 389 GTAAGTTEIVDSTG--LIHPAGEVAPLAKNIVRLASHAEQVSMGEKGYRYKEM 446
DB 306 ASKKGGISSEFVLDGITGYHLAEPMSSDSII---NDINRALADKERHQAERKSLVFSK 361
QY 447 FMEHMAERIAAVLKD 462
DB 362 YSMENVAQRFEQMK 377

Search completed: June 30, 2003, 16:30:26
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2003, 16:26:57 ; Search time 22 Seconds

(without alignments)
891.741 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2397
Sequence: 1 MAKTRSPFAVAAGRGPRVH.....ERIAVLKDVLRKSGHSRS 473

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	185.5	7.7	352	1	LPCC_RHILV
2	180.5	7.5	351	1	LPCC_RHILV
3	166.5	6.9	377	1	LPCC_RHILV
4	153	6.4	381	1	LPCC_RHILV
5	134.5	5.6	380	1	LPCC_RHILV
6	119.5	5.0	428	1	LPCC_RHILV
7	119.5	5.0	480	1	LPCC_RHILV
8	118	4.9	374	1	LPCC_RHILV
9	118	4.9	376	1	LPCC_RHILV
10	117	4.9	608	1	LPCC_RHILV
11	112.5	4.7	2564	1	LPCC_RHILV
12	108.5	4.5	605	1	LPCC_RHILV
13	107	4.5	340	1	LPCC_RHILV
14	107	4.5	603	1	LPCC_RHILV
15	107	4.5	862	1	LPCC_RHILV
16	106	4.4	358	1	LPCC_RHILV
17	105.5	4.4	508	1	LPCC_RHILV
18	105.5	4.4	808	1	LPCC_RHILV
19	104	4.3	778	1	LPCC_RHILV
20	104	4.3	805	1	LPCC_RHILV
21	103.5	4.3	608	1	LPCC_RHILV
22	103.5	4.3	807	1	LPCC_RHILV
23	102	4.3	353	1	LPCC_RHILV
24	101.5	4.2	536	1	LPCC_RHILV
25	101.5	4.2	607	1	LPCC_RHILV
26	100	4.2	615	1	LPCC_RHILV
27	100	4.2	801	1	LPCC_RHILV
28	100	4.2	1158	1	LPCC_RHILV
29	99.5	4.2	608	1	LPCC_RHILV
30	99	4.1	804	1	LPCC_RHILV
31	99	4.1	1056	1	LPCC_RHILV
32	98.5	4.1	805	1	LPCC_RHILV
33	98	4.1	377	1	LPCC_RHILV

34	98	4.1	512	1	MCPD_ENTAE
35	96.5	4.0	438	1	GLGA_THECA
36	96.5	4.0	409	1	C710_ARATH
37	96.5	4.0	609	1	UGST_ARATH
38	96.5	4.0	609	1	UGST_ORYZA
39	96.5	4.0	805	1	IP22_ORYZA
40	96.5	4.0	926	1	POOL_HAEIN
41	96	4.0	744	1	RELA_ECOLI
42	96	4.0	718	1	TRKA_ECOLI
43	96	4.0	808	1	SUSI_ORYZA
44	96	4.0	1057	1	SPSI_CITUN
45	95.5	4.0	643	1	HS70_CIAHE

ALIGNMENTS

RESULT 1	LPCC_RHILV	STANDARD;	PRT: 352 AA.
AC	068547;		
AD	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DE	Lipopolysaccharide core biosynthesis mannose 6-phosphate transferase (EC 2.4.1.16)		
GN	LPCC.		
OS	Rhizobium leguminosarum (biovar victoriae).		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
CC	Rhizobiaceae; Rhizobium.		
OK	NCBI_Taxid=387;		
RM	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3855;		
RX	MEDLINE=98434549; PubMed=9756877;		
RA	Kadmas J.L., Allaway D., Studholme R.E., Sullivan J.T., Ronson C.W.,		
RT	Pool P.S., Raetz C.R.H.;		
RL	"Cloning and overexpression of glycosyltransferases that generate the		
RU	lipopolysaccharide core of Rhizobium leguminosarum."		
CC	J. Biol. Chem. 273:126432-126440(1998).		
CC	-1- FUNCTION: ACTS AS TRANSFER OF MANNOSE GROUP TO A 3-DEOXY-D-MONO		
CC	OCTULONIC ACID (KDO) VIA AN ALPHA-1,5 LINKAGE.		
CC	-1- PATHWAY: Lipopolysaccharide core biosynthesis.		
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: AF050103; AAC0515.1;		
DR	InterPro: IPR001296; Glycosyltransferase.		
DR	Pfam: PF00534; Glycosyltransferase.		
DR	Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.		
SO	SEQUENCE 352 AA; 38640 MW; 8F3F5CE20C6CE08B CRC64;		
Query Match	7.7%; Score 185.5; DB 1; Length 352;		
Best Local Similarity	24.9%; Pred. No. 3.1e-07;		
Matches	94; Conservative 45; Mismatches 135; Indels 103; Gaps 18;		
QY	136 VQVLPARGQANDVIALKADVLTANTAVAGKMDPVYKDHVYKULPKLW-----		184
DB	26 VQVLPARGQANDVIALKADVLTANTAVAGKMDPVYKDHVYKULPKLW-----		66
QY	185 -----VHMRGKRYFKV-----EYKALP-----FVAGAMISHTAEYNSR-----T 223		
DB	67 RRRRWV-HARRNNEAVGILRLTLMPLKLPFSAQ-RRHATKMLIRMDAVIAT 124		
QY	224 SDRK-LQMPQYVYVHGNKELM-----EVAEDNARVRLREHRSIGVREDDLPFIIT 278		

Db 125 SDRSGSELEVPHTVIOH-GVDLALFHPPEAEDGIA-----ATGLPGRHLV-GCF 172

Qy 279 NSVSRGGDLPFOAFYQALQIHEKLVPRHVVVGSVDVNAQKTFEOLRDPVYKNT 338

Db 173 GRVHOKGTDLEVRMIELP--OHTEMT-----AVSGRTVAHYAFADKKADVAAG 225

Qy 339 IHDRVHFVNK-----TLAAVEYLAIDVLVONSQGRCEGRTITTEMAAFKLPV 387

Db 226 LSDRIIFLGEVDPDIKIMYRRLLTLVAVP-----SRNEGGLTPLEMAASRAV 272

Qy 368 LGTAAGGTETVLDSTGLHPAGKEGVAFLAKNIVRLASHAQRYSMGKGYGVXEMF 447

Db 273 VASAGAGYALVETGTGTSV-VAASSDGA-LTRALAPIADPALAAHGBNNLRVRANF 330

Qy 448 MEHMAERLAAVLKCVL 464

Db 331 ALERASAIQAVVNSIL 347

RESULT 2

LPSS RHIME STANDARD; PRT; 351 AA.

AC QGRSN2;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Lipopolysaccharide core biosynthesis mannosyltransferase lpsB

DE (EC 2.-.-.-)

GN LPSS OR R01572 OR SMC01219.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RCR2011 / S047;

RX MEDLINE=21101807; PubMed=1157937;

RA Lagree A., Hozbor D.F., Niehaus K., Pich Creer A.J.L., Lorenzen J.,

RA Arnold W., Puhler A.;

RT "Genetic characterization of a Sinorhizobium meliloti chromosomal

RT region involved in lipopolysaccharide biosynthesis.";

RL J. Bacteriol. 183:1248-1258(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,

RA Boistard P., Becker A., Bouctry M., Cadieu E., Dreano S., Gloux S.,

RA Godtke T., Goffeau A., Kahn D., Kiss E., Lelaute V., Masuy D.,

RA Pohl T., Portetelle D., Puhler A., Purnelle B., Ramperger U.,

RA Renard C., Thebault P., Vandebol M., Weidner S., Galibert F.;

RT "Analysis of the chromosome sequence of the legume symbiont

RT Sinorhizobium meliloti strain 1021.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).

CC -I- FUNCTION: ACTS AT TRANSFER OF MANNOSE GROUP TO A 3-DEOXY-D-MONO

CC OCTULONIC ACID (KDO) VIA AN ALPHA-1,5 LINKAGE (BY SIMILARITY).

CC -I- PATHWAY: Lipopolysaccharide core biosynthesis.

CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: AF193023; AAF06008.1; -

DR EMBL: AL591787; CAC6451.1; -

DR InterPro: IPR001296; Glycoe transf_1.

DR Pfam: PF00534; Glycoe transf_1; 1.

KM Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;

KM Complete proteome.

SQ SEQUENCE 351 AA; 38662 MW; 98436AB9DB82377A CRC64;

Query Match 7.5%; Score 180.5; DB 1; Length 351;

Best Local Similarity 24.3%; Pred. No. 7,7e-07;

Matches 84; Conservative 44; Mismatches 133; Indels 85; Gaps 14;

Qy 170 VLKDHVPKVLPEKTI-----LW-----WIENRCHYFRVEYVKGHPV----- 205

Db 41 VLDPGLPSPSLSVFRDLHLTWKREGRPCRW-HARRN-----VEMPAILLDLR 92

Qy 206 -----AGAMIDHTTAAYVNS-----RTSDRLK--IQMPOTVYVHAGNSKEIM 246

Db 93 MKRLVFTSASGRHHTG---MSKFLIRMDAVLATSGRTAAVLDPNVLHGLDITKRG 149

Qy 247 EVADENVARVLRHRESIGVRSDDLPAIINSVSRGKQDLFOAFYQALQIHEKLV 306

Db 150 PPEDKTEAKK-----ALGLDPKPKFVCGCFGRVRRHOKGTLPFDSMIALLPC----- 195

Qy 307 KVPRIAAVVGSDVNAQKTFEOLRDPVYKNTIHDRVHFVNKTLAAVEYLAIDVLVONS 366

Db 196 -RDPWGAIVAGRATGPHLAPESELKERVAKGLADRLIFVGHNTIIPMWALDLFV--A 252

Qy 367 QGRCEGRTITTEMAAFKLPVLTAAAGTTTIVLDGS--TGLHPAGK-----EGVAFLA 419

Db 253 PQRWEGFGLTPLEMAATGVVATDVGAFFSELVTCSEETGLIIAADLKVAVDAAPM 312

Qy 420 KNIVRLASHAQRYSMGKGYGVXEMFMEHMAERLAAVLKCVLR 465

Db 313 DDEPRLLA-----AASANGLARISKFNALIEQBRAIAVAYSSLWR 351

RESULT 3

CTSA BACSU STANDARD; PRT; 377 AA.

AC P46915;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Spore coat protein Sh.

GN CTSA.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / 60015;

RX MEDLINE=95400496; PubMed=7545510;

RA Abe A., Koide H., Kohno T., Matabe K.;

RT "A Bacillus subtilis spore coat polypeptide gene, cots.;"

RL Microbiology 141:1433-1442(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=9048467; PubMed=9387221;

RA Lapidus A., Gallenon N., Sorokin A., Ehrlich S.D.;

RT "Sequencing and functional annotation of the Bacillus subtilis genes

RT in the 200 Kb rrmh-dnaB region.";

RL Microbiology 143:3431-3441(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=9044033; PubMed=9384377;

RA Kunitz F., Ogasawara N., Moger I., Albertini A.M., Alloni G.,

RA Azevedo V., Bettler M.G., Beesler P., Boistard P., Borchert S.,

RA Borries R., Bouctry L., Brans A., Braun M., Briganti S.C., Bron S.,

RA Brouillet S., Bruch C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Ertan K.D., Erlingsson J., Fabre C., Ferrari E., Fougere D.,

RA Fritz C., Fujita M., Fuma Y., Galla Y., Galizzi A., Gallenon N.,

RA Ghin S.Y., Glaeser P., Goffeau A., Goughly E.J., Grandi G.,

RA Guisepi G., Guy B.U., Haga K., Haeck U., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lamber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara K., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemori K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpe P., Tognoni A.,
 RA Totsato V., Uchiyama S., Vandenhof M., Vannier F., Vassart A.,
 RA Viari A., Mamut R., Medler E., Medler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=99251148; PubMed=10234840;
 RA Takamatsu H., Kodama T., Watabe K.;
 RT "Assembly of the *CotA* coat protein into spores requires *CotS* in
 RT *Bacillus subtilis*.";
 RL FEMS Microbiol. Lett. 174:201-206(1999).
 CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D31847; BAA06633.1; -
 DR EMBL: AF008220; AAC00219.1; -
 DR EMBL: Z99119; CAB5069.1; -
 DR Subtilist; BG11381; cotSA.
 DR InterPro: IPR001296; Glycosyltransf.1.
 DR Pfam: PF00534; Glycosyltransf.1; 1.
 KW Sporulation; Transferase; Glycosyltransferase; Complete proteome.
 SQ SEQUENCE 377 AA; 42912 MW; 1F978E1B79F9E660 CRC64;
 CC -----
 Query Match 6.9%; Score 166.5; DB 1; Length 377;
 Best Local Similarity 24.6%; Pred. No. 11e-05; Indels 49; Gaps 18;
 Matches 85; Conservative 58; Mismatches 153; Indels 49; Gaps 18;
 Oy 134 HGQVLPARGGAVDIALKA---DVIINTVAGKWDVPLKDVHPKVLPRILMWHENR 190
 Db 58 HYVHLEDEYEAVGAEIKKSRFDIV-HVGNRPSWV-PLTKKQAPDAV-FILSVENEM 112
 Oy 191 GHYFKYEVKHLPRVAGAMIDS---HTTAEMNSRISDLKIQMPTVVVHNSKELM 246
 Db 113 FAYDKISQAE-----GECIDSVAGQVTVSDIGTITTSFPANSKTKVYSGVDIKY 167
 Oy 247 EVAEDNVARVLRHIESIGVSEDLPLALINSVSGKODFLQAFYALQDIQHEK 306
 Db 168 HPRWTEGQRA-REEMSEGLGKKIVL-FVGRISKVKRPHILLQALRDILE--EH--- 220
 Oy 307 KVRRIAAVVVGSDDVAQTKF--ETGLRDVVVNTIH-----DRNHPN--KLVAAPY 355
 Db 221 --PDVMMVFIS-----KMFQNDLNNV--KHLHLGAMQNDVTHFQVKKDIPRL 270
 Oy 356 IAAIDVLYVNSQGECEFGRIIT EAMAFKLPVLGTAAGTTEIVLDGST-LLHPAKEG 414
 Db 271 YTVSDVFCSSQMQ-EPLARVHYEAMAGLPITISNKGMPVEIEGKKGYIIDF--EN 327
 Oy 415 VAPLAKNIVTLASHEQGVSGKGVKGVKEMFNHMAIRIAY 459
 Db 328 PKQYARINDLSSSEKREKRGKYSREASFGQRYAENLISV 372

RESULT 4
 ID REPK_SALTY STANDARD; PRT; 391 AA.
 AC P26470;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lipopolysaccharide 1,2-N-acetylglucosaminetransferase (EC 2.4.1.56).
 OS REPK OR WAKK OR STM3714.
 GN Salmonella typhimurium.
 OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OK NCBI_TaxId=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=92041612; PubMed=1657881;
 RA MacLachlan P.R., Kadam S.K., Sanderson K.E.;
 RT "Cloning, characterization, and DNA sequence of the *rfaK* region for
 RT lipopolysaccharide synthesis in *Salmonella typhimurium* LT2.";
 RL J. Bacteriol. 173:7151-7163(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulyaney E.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -I- FUNCTION: ADDS THE TERMINAL N-ACETYL-D-GLUCOSAMINE GROUP ON THE
 CC GLUCOSE(II) GROUP OF LPS.
 CC -I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine +
 CC lipopolysaccharide = UDP + N-acetyl-D-glucosaminyl-
 CC lipopolysaccharide.
 CC -I- PATHWAY: Lipopolysaccharide core biosynthesis.
 CC -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE.
 CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -I- SIMILARITY: SOME, WITH B.SUBTILIS YTXN.
 CC -I- SIMILARITY: SHOWS VERY LITTLE SIMILARITY TO E.COLI REPK.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M73826; AAA27207.1; -
 DR EMBL: AB008672; AAL22573.1; -
 DR PIR: C41317; C41317.
 DR StyGene; SG10339; rfaK.
 DR InterPro: IPR001296; Glycosyltransf.1.
 DR Pfam: PF00534; Glycosyltransf.1; 1.
 KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 381 AA; 43152 MW; F60F37FF175372C6 CRC64;
 CC -----
 Query Match 6.4%; Score 153; DB 1; Length 381;
 Best Local Similarity 23.4%; Pred. No. 0.00012;
 Matches 74; Conservative 40; Mismatches 110; Indels 92; Gaps 11;
 Oy 168 DPLVDKHVPKVLPRILMWHENRGHYFVEVYVHLPVAGAMIDSHTAEMNSRISDL 227
 Db 133 EPELDNDNAKIL-----VPSQFLKAFYEERLPAAVSTIVPGCAETKYNRPDNL 183
 Oy 228 KIQMPTVVHNSKELMEVAEDNVARVLRHIESIGVSEDLPLALINSVSGKQ 287

Db 184 RQO-----LNTADPATV-----LLYA--GRISPDGI 208
Qy 288 DLFIOAFYQALOLIOHEKXKVPRIHAWVSGDVNAQTEKETOJDPVVKXTHIDRHVHN 347
Db 209 LLLLOAFQJQRTLRSHIKL-----VVQG-DPYASRGE-----FAEYQK 246
Qy 348 KTLAVAPYLA-----IDLVYVNSQGRGECFGRITTEAMAFELPVL 388
Db 247 KVLDAAEKIGTDCIMAGSQSPDMHNFYHADIIVPSQVE-EAFQVAVAEAMAAKAVL 305
Qy 389 GTAAGTTEIVLDGSTG--LHPAGKEGVAFLAKNIVRLASHAQRVSGEGYGRVEM 446
Db 306 ASKKGISSEFVLDDIGTGHIAEPMSDSII-----NDIRALADERQIENAKSLVFSK 361
Qy 447 FWEHMAERIAVLKQ 462
Db 362 YSMENVAQRPEQMK 377

RESULT 5

CAPM STRAU STANDARD; PRT; 380 AA.
AC P39862;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Capsular polysaccharide biosynthesis glycosyl transferase capM
DE (EC 2.4.1.16).
GN CAPM.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=M;
RX MEDLINE=95050273; PubMed=7961465;
RA Lin W.S., Cunneen T., Lee C.Y.;
RT "Sequence analysis and molecular characterization of genes required
RT for the biosynthesis of type 1 capsular polysaccharide in
RT Staphylococcus aureus";
RL J. Bacteriol. 176:7005-7016(1994).
CC -1- FUNCTION: REQUIRED FOR THE BIOSYNTHESIS OF TYPE 1 CAPSULAR
CC POLYSACCHARIDE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U10927; AAA64652.1; -
DR InterPro: IPR001296; Glycosyl transf_1.
DR Pfam: PF00534; Glycosyl transf_1.1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 380 AA; 43166 MW; 0264E8376CD1A9EF CRC64;

Query Match 5.6%; Score 134.5; DB 1; Length 380;
Best Local Similarity 25.3%; Pred. No. 0.0035;
Matches 46; Conservative 38; Mismatches 69; Indels 29; Gaps 6;
Qy 299 QLIQHEKTVPR---IHAVVSGDVNAQTK-----FTQJRDVPAVKTIHRRVHF 346
Db 214 ELIGSFILVSGYNVKKLVIGS---LETNSIDSDYLFETQNPVNVILKIVSDPSISFY 270
Qy 347 NKTIAVAPYLAIDVLYVNSQGRGECFGRITTEAMAFELPVLGTAAGTTEIVLDGSTGL 406
Db 271 NN-----MNVFVFPTRHG--FGNVSEIQAQLEVPITTNVGAIDIVVNGGEFG 318
Qy 407 LHPAGKEGVAFLAKNIVRLASHAQRVSGEGYGRVEMHMAERIAVLKQVLK 466

Db 319 I--VEKGDPRKAIKAEIKELTNDSELAETIGNKRKRVENKFSQIIWELBSMNTFLKE 376
Qy 467 SQ 468
Db 377 SE 378

RESULT 6

Y486 MYCLE STANDARD; PRT; 428 AA.
ID Y486 MYCLE
AC P54138; Q9CSB50.
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ML2443.
GN ML2443 OR U2168F OR B2168_C2_201.
GN Mycobacterium leprae.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RP
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Dutfoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO M.TUBERCULOSIS RV0486.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U00018; AAA17228.1; ALT_INIT.
DR EMBL: AL583925; CAC31960.1; -
DR Leproma: ML2443; -
DR InterPro: IPR001296; Glycosyl transf_1.
DR Pfam: PF00534; Glycosyl transf_1.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 428 AA; 45291 MW; A14F9F0187E3587C CRC64;

Query Match 5.0%; Score 119.5; DB 1; Length 428;
Best Local Similarity 21.0%; Pred. No. 0.063;
Matches 90; Conservative 53; Mismatches 174; Indels 111; Gaps 16;
Qy 54 GRDPRALNTVAVGSPGLFMSKVLVLVSHSLSGPFLMELAFILRVNSQVWYTT 113
Db 13 GIGDAGKNAVYVLQ-SALHLARGIEVEIFTRATASADP-----PIVWA 56
Qy 114 NQROETNDVTVSLDEHMLNHGVYLPARGQAVDI-----ALKADVLINTAVAGKMD 168
Db 57 -----PGVLVNVVAGPPEGIDKYDLPYQCAFAAGVLRARAAHEPQYD 101
Qy 169 PVLKDVPRVLPKILMIEHMGHFKVEYVXHLPRVAGAMDSHTTAEVNSASTSRUK 228
Db 102 IVHSH-----YVLSGQVGMILARDRA--VPLVHTA-----HTLAAVKNAALADGDA 145

QY 229 ICMPTVY--VHLGNSKEIMEVADNVARVLRH----- 261
Db 146 APPPLRSVGEQVVDENMIVTDDERKOLISIHADPAKIDVHNPGLDMFRGDR 205
QY 262 -TRESLGVRSDDLPAIINSVSRGQDLFLQAFYQALQIHEKLVKPRIHAVVGSV 320
Db 206 AARALGIPLDGNVAVFVGRIQPLAAPDIVRA-----AAKLPOVRIYVAGPS 254
QY 321 NQOTFEFQLDFVVKNTIHDRVHFV--NKTIAVAPYLAIDVLVWNSQGRGCEGRIT 377
Db 255 GGGIASPGLVRLADEIGTAVTTLPPQSTINATVFOADLVAVPS--YSESGIYA 311
QY 378 IEAMAPKLPVLGTAAGTTEIYLDGSGTL--HPAG-----KEGAPLAKNIYR 424
Db 312 VERKOCCTPVVAAVAGGLPVAVADGVTGLVFGHNVGMADVQDLRLSLAGPOARISR 371
QY 425 LA-SHAQ 431
Db 372 AAVVHBAQ 379

RESULT 7
ID Y486 MYCTU STANDARD; PRT; 480 AA.
AC 011152;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Rv0486.
GN Rv0486 OR MT0504 OR MTCY20G9.12.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jorgensen K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Cwim M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Weidman J.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weisman J., Kouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO M.LBRAR ML2443.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL: 277162; CAB09947.1; -
DR EMBL: AE006951; AAK44727.1; -.

DR TIGR; MT0504; -
DR TubercuList; Rv0486; -
DR InterPro; IPR001296; Glycos transf_1.
DR Pfam; PF00534; Glycos transf_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 480 AA; 50541 MW; 2134755E894A9CCF CRC64;

Query Match 5.0%; Score 119; DB 1; Length 480;
Best local similarity 20.8%; Pred. No. 0.081;
Matches 95; Conservative 61; Mismatches 178; Indels 122; Gaps 18;

QY 47 ALRDECDGRC-----DPAALNTAASGSPPLGFWRSKVLVLSHSLIS-----G 90
Db 14 AARRVYRGEGATSRSGPSGNSRNVSAADP-----RRVALLVHTSPLAOGTGDAGG 67
QY 91 GGLIMELAFLLHNVGSQVWITNORSGEINDVYTSLSHRMLNNGVQVPAHQGEAVDIA 150
Db 68 MNVYMLQSLHLARRGIEVEIFPTATASADPV-----VAVAG----- 106
QY 151 LKADLVILNTAVAGKW--LDPVLDKDHVPKVLPKILMW-----IHE-----MRGHYF-- 194
Db 107 -----VLTVNVVAGPEGLD--KYDLPTQLCAFAGVLRABAVHEPGYDVIHSHYMLS 158
QY 195 -KYEVYKLPVAVAGAMIDSHTTAEYNSRTSDRLKIQMPTVYVHLGNSKEIME-----V 248
Db 159 GGQWGLARDAVAVPLVHTAHTLAAYKNAALADG--DGEPLRTVGEQVDEADRLIV 215
Db 216 NNDDPARGVSLHGADPARIDVHGVLDVFRPGRBARAALGLPVDERVAVAFVGRIO 275
QY 249 AEDNVAARVLRH-----IRESLGVRSDDLPAIINSVS 282
Db 283 RKGQDLFLQAFYQALQIHEKLVKPRIHAVVGSVDVAQKTEQLRPFVVKNTIHR 342
QY 276 PKKADIVLRA-----AAALPEVRIYVAGPSGSLASPDGLVRLADEIGISR 324
Db 343 VHFV--NKTIAVAPYLAIDVLVWNSQGRGCEGRITTEBMAFKLPVLGTAAGTTEIYL 400
Db 325 VTELPQSHDTLATLFRADLVAVPSY--SESGIYAVAGQCTPVVAAVAGLPAVR 382
QY 401 DSGTGLL--HPAGKGVAPLAKNIVRLAASHAQRV 433
Db 383 DGTITLVSGHEVQ--WDADIDHLRLCGPGRV 416

RESULT 8
ID RFAG_ECOLI STANDARD; PRT; 374 AA.
AC P25740;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide core biosynthesis protein rfaG
DE (Glucosyltransferase I).
GN RFAG OR WAAG OR PCSA OR B3631.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteriia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9211133; PubMed=1732225;
RA Parker C.T., Pradel E., Schnaitman C.A.;
RT "Identification and sequences of the lipopolysaccharide core
RT biosynthetic genes rfaG, rfaF, and rfaG of Escherichia coli K-12."
RL J. Bacteriol. 174:930-934 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RC Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."

RL Nucleic Acids Res. 22:2576-2586(1994).

RN [3]

RP SEQUENCE OF 1-217 FROM N.A.

RX MEDLINE=93077458; PubMed=1447141;

RA Clementz T.;

RT "The gene coding for 3-deoxy-manno-octulosonic acid transferase and

RT the rfaG gene are transcribed from divergently arranged promoters in

RT *Escherichia coli*."

RL J. Bacteriol. 174:7750-7756(1992).

RN [4]

RP SEQUENCE OF 1-58 FROM N.A.

RX MEDLINE=95080611; PubMed=7988890;

RA Filipiniec E., Huisman T.T., Willemsen P.T., Appelmeik B.J.,

RA Graaf F.K., Oudega B.;

RT "Identification by Tn10 transposon mutagenesis of host factors

RT involved in the biosynthesis of K99 fimbriae of *Escherichia coli*;

RT effect of LPS core mutations."

RL FEMS Microbiol. Lett. 123:201-206(1994).

CC -1- FUNCTION: INVOLVED IN THE ADDITION OF THE FIRST GLUCOSE RESIDUE

CC TO THE LIPOLYSACCHARIDE CORE.

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M80599; AAA24082.1; -

DR EMBL; M86305; AAA03743.2; -

DR EMBL; U00039; AAB1608.1; -

DR EMBL; AE000440; AAC76655.1; -

DR EMBL; S75736; AAD45826.1; -

DR PIR; B42595; B42595.

DR PIR; S27559; S27559.

DR EcoGene; EG1339; rfaG.

DR InterPro; IPR001296; Glycosyltransf_1.

DR Pfam; PF00534; Glycosyltransf_1; 1.

KM Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;

KM Complete proteome.

SO SEQUENCE 374 AA; 42284 MW; 7720AE577CEB80C2 CRC64;

Query Match 4.9%; Score 118; DB 1; Length 374;

Best Local Similarity 24.2%; Pred. No. 0.069;

Matches 50; Conservative 40; Mismatches 91; Indels 26; Gaps 9;

QY 259 REHRESLGVRSDELFAINSVSRGQDLFLQAFYQALQIHEKLVPRHAIVVGS 318

DB 103 REIRONGKIKEDQNTLQVSDGRKGVDRSTELASLPESLRHNTL-----LFVVGQ 236

QY 319 DVNAQYFEETQLRDVFNKNTIDRVHFNKTLAVAPYLAIDVIVNOSGRCGECFGRIT 378

DB 237 D--KPRFEA---LAETKLGVRNVHFFSGNDVSELMMAADLLHPAV--GEAGIVTL 288

QY 379 EAMAFKLPVGTAAAGTEIVLDGSG--LHPACKREGVAAIAKNIVRLA-SHAQGRISM 435

DB 289 EAIKAGLPVLTNAGCVAHYIADANGCTVIAPFSQGL---NEVLRALTLQSPLRVAV 344

QY 436 GE--KGYGRVKEFMFHHMAERIAVAVL 460

DB 345 AENNAHYADTDLY---SLPEKADII 368

RESULT 9

ID VIPC_SALTI STANDARD; PRT; 578 AA.

AC 004975.

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Vi polysaccharide biosynthesis protein vipc/tyie.

GN VIPC OR TYIE OR STY4656.

OS Salmonella typhi.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OK NCBI_TaxID=601;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TY2;

RX MEDLINE=94069051; PubMed=8248629;

RA Maxin H., Virlogeux I., Kolyva S., Popoff M.Y.;

RT "Identification of six open reading frames in the *Salmonella enterica*

RT subsp. *enterica* ser. Typhi viab locus involved in Vi antigen

RT production."

RL Res. Microbiol. 144:363-371(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=GIFU 10007;

RX MEDLINE=9332324; PubMed=8331073;

RA Hashimoto Y., Ii N., Yokoyama H., Ezaki T.;

RT "Complete nucleotide sequence and molecular characterization of viab

RT region encoding Vi antigen in *Salmonella typhi*."

RL J. Bacteriol. 175:4456-4465(1993).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,

RA Krogh A., Larsen T.S., Leach S., Moulé S., O'Gara P., Parry C.,

RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,

RA Whitehead S., Barrett B.G.;

RT "Complete genome sequence of a multiple drug resistant *Salmonella*

RT *enterica* serovar Typhi CT18."

RL Nature 413:848-852(2001).

CC -1- PATHWAY: Vi polysaccharide biosynthesis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X67785; CAA37994.1; -

DR EMBL; D14156; BAA03195.1; -

DR EMBL; AL627283; CAD06776.1; -

DR InterPro; IPR001296; Glycosyltransf_1.

DR Pfam; PF00534; Glycosyltransf_1; 1.

KM Complete proteome.

FT CONFLICT 290 S -> G (IN REF. 1).

FT CONFLICT 362 KH -> ND (IN REF. 1).

FT CONFLICT 363

SO SEQUENCE 578 AA; 65009 MW; 8D420563D868C189 CRC64;

Query Match 4.9%; Score 118; DB 1; Length 578;

Best Local Similarity 22.2%; Pred. No. 0.13;

Matches 69; Conservative 58; Mismatches 132; Indels 52; Gaps 14;

QY 155 LVTLNNAVAGMDPVLAKDHPV-----LPKILWIMHEMGHFKVEY---VKHLPVY 205

DB 294 WMLMALING-----VPRILQGLRGILPPV-----VKKLFFKEPEPYQALAVV 337

QY 206 AGA--MDSHTTAETVNSRSDPLKIQMPQTVVYHLSGSKELMEVAEDNVAARVLEHIR 263

DB 338 PGVDEFSNNHCVTRHY---ADWLKLEAKHPQYVYNGVLPSPTESSS--VPHKIMQGFQ 392

QY 264 ESIQGRSDDLFAINSVSRGQDLFLQAFYQALQIHEKLVPRHAIVVGS-SDVNA 322

DB 393 KI---QDAD---TTIGVFRFVGVGDKNPFAMIDPARYLQHH---PATRFVVGDDDLRA 442

323 QTEFETOLDRBVPVKNTHRVFVVKTLAVAPYLAIDVWNSQGRGECGRITIEPAMA 382
 Db 443 EAOGRAE-----QGLIERILIFVASRSDVGYLQKNVFLTS--RYGGLPNTLIDQM 494
 QY 383 FKLPVLGTRAGGTTEYLDGSGT-LHPAGKEGVAPLANKIVRLASHAQRVSKEKGKG 441
 Db 495 VGVPLVSTPAPGSAECRIGVSGFTLDDAQTVALDQACRAEKVLVLMRSRTGICDOOTS 554
 QY 442 RVKEMFEHEHM 452
 Db 555 FLDERTVEHM 565
 RESULT 10
 UGST MANES STANDARD; PRT, 608 AA.
 ID UGST MANES STANDARD; PRT, 608 AA.
 AC Q43784;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Granule-bound glycogen [starch] synthase, chloroplast precursor
 (EC 2.4.1.11).
 GN WAKY OR GBSS.
 OS Manihot esculenta (Cassava) (Manioc).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 SC Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots; Rosidae;
 OC eurosida I; Malpighiales; Euphorbiaceae; Manihot.
 OX NCBI_TaxID=3983;
 RN [1]
 RP SEQUENCE FROM N A.
 RC STRAIN=CV. M.COL 22, TISSUE=Tuberous root;
 RX MEDLINE=94083565; PubMed=8260633;
 RA Salehuzzaman S.N., Jacobsen E., Visser R.G.F.;
 RT "Isolation and characterization of a cDNA encoding granule-bound
 starch synthase in cassava (Manihot esculenta Crantz) and its
 RT antisense expression in potato."
 RL Plant Mol. Biol. 23:947-962(1993).
 CC - FUNCTION: RESPONSIBLE FOR THE SYNTHESIS OF AMYLOSE IN RESERVE
 STARCH.
 CC - CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-alpha-D-glucosyl} (N) =
 CC UDP + {(1,4)-alpha-D-glucosyl} (N+1).
 CC - PATHWAY: Starch biosynthesis.
 CC - SUBCELLULAR LOCATION: CHLOROPLAST, GRANULE-BOUND.
 CC - TISSUE SPECIFICITY: SYNTHESIZED IN A NUMBER OF DIFFERENT ORGANS,
 CC BUT MOST ABUNDANTLY IN TUBERS.
 CC - SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE
 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL: X74160; CAA52273.1; -
 DR InterPro: IPR0001296; Glycosyltransferase_1.
 DR Pfam: PF00534; Glycosyltransferase_1.
 KW Glycosyltransferase; Transferase; Glycosyltransferase;
 KW Transglutaminase; Chloroplast; Starch biosynthesis.
 FT TRANSIT 1 78 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 79 608 GRANULE-BOUND GLYCOGEN [STARCH] SYNTHASE.
 FT BLINDING 96 96 UDP-GLUCOSE (BY SIMILARITY).
 SO SEQUENCE 608 AA; 66968 MW; C9C97C0D311BDBD CRC64;
 Query Match 4.9%; Score 117; DB 1; Length 608;
 Best Local Similarity 22.6%; Pred. No. 0.16; Indels 64; Gaps 11
 Matches 56; Conservative 37; Mismatches 91;

```

Cc Db      359  KNPVYDKTIDIHVDATITVMD----- AKDLKEALQAEVGLPDSVNPDLIG 403
Cc Oy      277  IINSRSRGGQDFLFOAFQALQLOIEBKAKPRIRHANVVGSDVNAQTKFETQIDPFVVK 336
Cc Db      404  FIFRLEEKSGSDIFVAAL--SOLVEH-----NQVITIST--GKKKFEKQTE----- 446
Cc Oy      337  NTHIDRVHVFVKTLAAVAPY-----LALIDVLVONSOGKGCFCGRITTEMAFKPLVL 388
Cc Db      447  --HLEVLVYDKRAGVAKENPVLAMHTAGDPMFLVPS--RFEPCGLIOLHAMRYGVPI 501
Cc Oy      389  GTAAGGTETIYLDGSGGL-----LH-----PAGEGVIAPLAKIYRLASHAEQVSMGKGY 440
Cc Db      502  VASTGGSLVDVITKSGYGVFGPMGALHVECDKIDSADVAALVTKVAR-----ALGTYAT 552
Cc Oy      441  GRVKEWFM 448
Cc Db      553  ALREWIL 560

Cc RESULT 11.
Cc SPEC HUMAN
Cc ID SPEC HUMAN STANDARD; PRT; 2564 AA.
Cc AC 09H254; 09HCD0; 09H3G8; 09HK17; 09HK18; 09HK19;
Cc DT 15-JUN-2002 (Rel. 41, Created)
Cc DT 15-JUN-2002 (Rel. 41, Last sequence update)
Cc DT 15-JUN-2002 (Rel. 41, Last annotation update)
Cc DE Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3)
Cc DE (Beta-IV spectrin).
Cc GN SPTBN4 OR SPTBN3 OR KIAA1642.
Cc OS Homo sapiens (Human).
Cc OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Cc OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Cc OX NCBI_Taxid=9606;
Cc RN [1]
Cc RP SEQUENCE FROM N.A. (ISOFORM 1).
Cc RX MEDLINE=21316449; PubMed=11294830;
Cc RA Tse W.T., Tang J., Jin O., Kozegren C., John K.M., Kung A.L.,
Cc RA Gwynn B., Peters L.L., Lux S.E.;
Cc RT "A new spectrin, beta-IV, has a major truncated isoform that
Cc RT associates with promyelocytic leukemia protein nuclear bodies and the
Cc RT nuclear matrix.";
Cc RT J. Biol. Chem. 276:23974-23985 (2001).
Cc RN [2]
Cc RP SEQUENCE FROM N.A. (ISOFORM 1 TO 4).
Cc RX MEDLINE=20539976; PubMed=11086001;
Cc RA Berghs S., Aguilero D., Dirix R. Jr., Maksimova E., Stabach P.,
Cc RA Herneel J.-M., Zhang J.-P., Philbrick W., Slepnev V., Ott T.,
Cc RA Solimena M.;
Cc RT "BetaIV spectrin, a new spectrin localized at axon initial segments
Cc RT and nodes of ranvier in the central and peripheral nervous system.";
Cc RT J. Cell Biol. 151:985-1002(2000).
Cc RN [3]
Cc RP SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).
Cc RC TISSUE=Brain;
Cc RX MEDLINE=20450683; PubMed=10997877;
Cc RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
Cc RT "Prediction of the coding sequences of unidentified human genes.
Cc RT XVIII. The complete sequences of 100 new cDNA clones from brain which
Cc RT code for large proteins in vitro.";
Cc RL DNA Res. 7:373-381(2000).
Cc CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1 (shown here), 2, 3 and 4; are
Cc CC produced by alternative splicing.
Cc CC -1- TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic
Cc CC islets.
Cc CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
Cc CC -1- SIMILARITY: CONTAINS 2 CALPOININ-HOMOLOGY (CH) DOMAINS.
Cc CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
Cc CC -1- SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.
Cc CC -----
Cc CC This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc CC between the Swiss Institute of Bioinformatics and the EMBL outstation
Cc CC at the European Bioinformatics Institute. There are no restrictions on its
Cc CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF311855; AAC42473.1; -
 DR EMBL; AF082075; AAC38874.1; -
 DR EMBL; AY004226; AAF93171.1; -
 DR EMBL; AY004226; AAF93172.1; -
 DR EMBL; AY004227; AAF93173.1; -
 DR EMBL; AB046862; BAB13468.1; -
 DR HSSP; 001082; 1BRK.
 DR Genew; HGNC:14896; SPTBN4.
 DR MIM; 606214; -
 DR InterPro; IPR001589; Actbind_actin.
 DR InterPro; IPR001715; Calpoin-like.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001605; Spectrin_PH.
 DR Pfam; PF00307; PH; 3.
 DR Pfam; PF00435; Spectrin; 56.
 DR PRINTS; PR00683; SPECTRINPH.
 DR SMART; SM00033; CH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00150; SPEC; 16.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS50021; CH; 2.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
 KW Alternative splicing.
 FT DOMAIN 1 202 ACTIN-BINDING (BY SIMILARITY).
 FT 1 165 CH 1.
 FT DOMAIN 180 282 CH 2.
 FT REPEAT 309 354 SPECTRIN 1.
 FT REPEAT 398 419 SPECTRIN 2.
 FT REPEAT 429 533 SPECTRIN 3.
 FT REPEAT 535 642 SPECTRIN 4.
 FT REPEAT 644 771 SPECTRIN 5.
 FT REPEAT 773 879 SPECTRIN 6.
 FT REPEAT 881 985 SPECTRIN 7.
 FT REPEAT 1019 1086 SPECTRIN 8.
 FT REPEAT 1088 1197 SPECTRIN 9.
 FT REPEAT 1199 1303 SPECTRIN 10.
 FT REPEAT 1305 1408 SPECTRIN 11.
 FT REPEAT 1410 1513 SPECTRIN 12.
 FT REPEAT 1515 1619 SPECTRIN 13.
 FT REPEAT 1621 1725 SPECTRIN 14.
 FT REPEAT 1727 1832 SPECTRIN 15.
 FT REPEAT 1834 1940 SPECTRIN 16.
 FT REPEAT 1942 2046 SPECTRIN 17.
 FT REPEAT 2048 2107 SPECTRIN 18.
 FT DOMAIN 2418 2527 PH.
 FT 1 1257 MISSING (IN ISOFORM 3).
 FT VARSPLIC 1258 1286 AV0A8GLRACGNYGRGQAVYRLKK -> MHPYSCS
 FT VARSPLIC 1287 1309 SABSCTPPPTQQLKRRR (IN ISOFORM 3).
 FT VARSPLIC 1310 2564 NOEMLRQOMQKLDLELOH -> CLTHPALHPWE
 FT VARSPLIC 2113 2154 PVIYPRSS (IN ISOFORM 2).
 FT 1 MISSING (IN ISOFORM 2).
 FT 1 IEKIKAESKQPTPLGRKFGPPTELAKAPILRPGY
 FT 1 E -> PREDELNPGVDPQMPQHTKRPSPKTKANKETAR
 FT 1 RDGTL (IN ISOFORM 4).
 FT 1 MISSING (IN ISOFORM 4).
 FT 1 MISSING (IN REF. 2).
 FT 1 L -> S (IN REF. 2).
 FT 1 E -> K (IN REF. 2).
 FT 1 E -> K (IN REF. 2).
 FT 1 G -> S (IN REF. 1).
 FT 1 CONFLICT 1331 1331
 FT 1 SEQUENCE 2564 AA; 288982 MW; 52CDE7D1D601EC6C CRC64;

Query Match 4.7%; Score 112.5; DB 1; Length 2564;
 Best Local Similarity 21.2%; Pred. No. 2.7; Indels 149; Gaps 21;
 Matches 95; Conservative 58; Mismatches 147;

QY 8 AYAAVAGRGVPHNRQTQLLLLLLVAAVASASTAGFL- -RGALRDCPGRGDPALNTA- 64
 DB 1107 AQEAAAGSGGFLPN-----SLEEDALLARAAKEVDQREEDARIVAA 1152
 QY 65 -----VAGSGPLG-----EMRSKI 78
 DB 1153 SEALLAADAELEGQALDDEMLPHLELGMWKEARREALVQAHTYQLFLNDLQAL 1212
 QY 79 VLVSHSLISGGPL-----LMELEFLRHVGSQVWITNORSQSTNVTYSLE 128
 DB 1213 VVLRNDEMALSDELPGTVSEVBAKQHRDPL-----TMEISQCKQVAVQA 1262
 QY 129 HRLMNGVOVLPARGQAVDIALKADVILNTAVAGKL-----DPYLKDHVPKYLPITIM 184
 DB 1263 EGLLRQG-NVGEQAQAEVTRLEKQOE--NQLRAQOMQKLDLELQHLRDCHLDG 1319
 QY 185 WHEMNGHFKVEYVKLPVAGAMIDSHTTAEYWNERTSDRLKIOMPQTVVHLGNSKE 244
 DB 1320 WHE-----KLMARDSTREDNHKHKWLH-----QAFMELANKE 1358
 QY 245 LMEVADNVARRVLR-----HIRESG-VRESDLLPAINSVRGQDLF----- 290
 DB 1359 WLEKIE-REGQLOMEKPELAAVSRKUGETRO--CWMELESTTOAARQLFPAASDAQ 1414
 QY 291 -LQAFYQALQLOHEKLVPRHAAVVGSD--VNAQTK-----PFTQLRDPVVKTIHR 342
 DB 1415 LVGSFAELDKKLHME--SQLODVPDGGDLATVNSQLKLSMESQVEEM-----YRE 1465
 QY 343 VHVVNKTLAVAPYLAIDIVY--QNSQG 368
 DB 1466 VGELOAQTLALPLEPASKELVGERQAVG 1494

RESULT 12
 UGST MAIZE
 ID UGST MAIZE STANDARD: PRT; 605 AA.
 AC P04713.
 DT 13-AUG-1987 (Rel. 05, Last Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Granule-bound glycogen [starch] synthase, chloroplast precursor
 DE (EC 2.4.1.11).
 GN MAXY.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kloeber R.B., Gierl A., Schwarz-Sommer Z., Saedler H.;
 RT "Molecular analysis of the waxy locus of Zea mays";
 RL Mol. Genet. 203:237-244(1986).
 CC -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AMYLOSE IN ENDOSPERM.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4)-alpha-D-glucosyl)(n) =
 CC UDP + ((1,4)-alpha-D-glucosyl)(n+1).
 CC -1- PATHWAY: Starch biosynthesis.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, GRANULE-BOUND.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X03935; CA27574.1; -
 DR EMBL; M24256; AAA33520.1; -
 DR PIR; S07314; S07314.

DR MaizeDB; 15806; -, Glycosyltransf. 1.
 DR InterPro; IPR001296; Glycosyltransf. 1.
 DR Pfam; PF00534; Glycosyltransf. 1; 1_-
 KW Glycosyl biosynthesis; Transferase; Glycosyltransferase;
 KW Transit peptide; Chloroplast; Starch biosynthesis.
 FT TRANSIT 1 72 CHLOROPLAST.
 FT GRANIN 73 605 GRANULE-BOUND GLYCAGEN [STARCH] SYNTHASE.
 FT BINDING 91 91 UDP-GLUCOSE (BY SIMILARITY).
 SQ SEQUENCE 605 AA; 65966 MW; 137F15207DFEC189 CRC64;

Query Match 4.5%; Score 108.5; DB 1; Length 605;
 Best Local Similarity 22.8%; Pred. No. 0.75; Mismatches 73; Indels 41; Gaps 7;
 Matches 44; Conservative 35;

QY 254 ARRVLRREHRSIG--VRESDLLFAIINSVRGGDLEQAFYALQLOHEKLVKPRI 311
 DB 374 AKAKMKALQAEVGLPVDRNIPLVAFIGRLERQKGPVMAAAIPQIMEMVED-----V 426
 QY 312 HAVVWSVDMNAQTFFETQLDFVVKNTLHDFVFNKTLAVAPYLAAL-----DVLV 363
 DB 427 QIVLIGT---GKKTFEKMLSAEK-----PRGVRAVVKFMAALAHIMAGDVLA 475
 QY 364 QNSQGRGCFGRITTEMAFKLPVLGTAGTTEIVLDGST-----LHPACK 412
 DB 476 VTS--RFPFCGLQLOGMRYGTPCACASTGLVDTIIEGKTGFHMGRLSDVCNVEPADV 533
 QY 413 EGVAFLAKNIVRL 425
 DB 534 KKVATTLQRAIKV 546

RESULT 13
 LPSE RHIME STANDARD; PRT; 340 AA.
 ID LPSE RHIME
 AC Q9R9N1;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lipopolysaccharide core biosynthesis glycosyl transferase lpse
 DE (EC 2.-.-.-).
 GN LPSE OR R01571 OR SMC01220.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RCR2011 / S047;
 RX MEDLINE=2136507; PubMed=11481430;
 RA Lagares A., Hozbor D.F., Niehaus K., Pich Otero A.U.L., Lorenzen J.,
 RA Arnold W., Puehler A.;
 RA "Genetic characterization of a Sinorhizobium meliloti chromosomal
 RT region involved in lipopolysaccharide biosynthesis";
 RT J. Bacteriol. 183:1248-1258(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=2136507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreaano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kils E., Lelaune V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ransberger U.,
 RA Renard C., Thebaud P., Vandenbol W., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF193023; AAF06009.1; -;
 DR InterPro; IPR001296; Glycosyltransf. 1.
 DR Pfam; PF00534; Glycosyltransf. 1; 1_-
 KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 340 AA; 38036 MW; D8363F95FD9E169F CRC64;

Query Match 4.5%; Score 107; DB 1; Length 340;
 Best Local Similarity 19.8%; Pred. No. 0.44; Mismatches 140; Indels 114; Gaps 17;
 Matches 76; Conservative 53;

QY 130 RMLNNGVQLPARQEAVIDA-----LKADLVLTAVAGKW 166
 DB 16 RFPVHLYNALAEVQQTALIRPGGRDRDIEGAKIRSHFRMLSDRIIL----- 67
 QY 167 LDPVLKQHV-----PKVLKILMW---IHEMRGHY---FKVEYVGHLPVAGAMIDSHTT 215
 DB 68 --PLKVKHMAHREKRPVL--MAMAPASLMPNYKGAFKISRLGDYF-----TR 112
 QY 216 AEYVNSRTSRLKIQMP--QTYVHLGNSKELMEVAEDNVARVLRHREISLGVASEDL 273
 DB 113 LSYF--RNTDCIVCNTPGIAERVS DLMKREIRVISNFTGTGRVY----- 155
 QY 274 LPAIINSVSRGK-----GODLFLQAFYALQLOHEKLVKPRIHAVVWSGDV 320
 DB 156 -----ANDRAKLDPADAPVVMGKRFERKGFHTLBAV----ALPQVYIMLIDGE 205
 QY 321 NAQT--KEETQLRDPVVKNTLHDFVFNKTLAVAPYLAALDVLVNSQGRGCFGRITTI 378
 DB 206 ERDNLHKLATDL-----GVSGRVFPAQMDTPRFLAAVDVVMSS--HEPLGANVLL 256
 QY 379 EAMAFKLPVLGTAGTTEIVLDGSTGLHPAG--KEGVAPLAKNIIVLASAQRVSMGE 437
 DB 257 ESMAGTPTVSTSBEPQMFHGDGKNGMLVDTIDBAG--FARAIEQIVADNSLRILAE 313
 QY 438 KQYGRVXEMFMHHMAERIALAVL 460
 DB 314 RGHETLVQSFSSRAITDAVLOLL 336

RESULT 14
 UGST HORVU STANDARD; PRT; 603 AA.
 ID UGST HORVU
 AC P05842;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Granule-bound glycogen [starch] synthase, chloroplast precursor
 DE (EC 2.4.1.11).
 GN WAXY.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Scrophophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 OC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv Vogelzanger Gold; TISSUE=leaf;
 RX MEDLINE=8830345; PubMed=2970062;
 RA Rhode W., Becker D., Salami F.;
 RA "Structural analysis of the waxy locus from Hordeum vulgare";
 RL Nucleic Acids Res. 16:7185-7186(1988).
 RN [2]
 RP SEQUENCE OF 76-89.
 RC STRAIN=cv, H354-295-2-5; TISSUE=Starchy endosperm;
 RX MEDLINE=94170739; PubMed=8125056;
 RA Flengserud R.;
 RT "Separation of acidic barley endosperm proteins by two-dimensional

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2003, 16:18:51 ; Search time 36 Seconds
(Without alignments)
2707.230 Million cell updates/sec

Title: US-09-938-294-45
Perfect score: 2397
Sequence: 1 MAKTPSFAVAAGKGPVH.....ERIAVLKDVAKSKQHSRS 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_protein:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriophage:*
 - 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1478	61.7	402	10 Q9FWT0	Q9FWT0 Arabidopsis
2	1471	61.4	458	10 Q9FXG9	Q9FXG9 Arabidopsis
3	671	28.0	188	10 Q9L148	Q9L148 Oryza sativa
4	487	20.3	697	10 Q9LSB5	Q9LSB5 Arabidopsis
5	482.5	20.1	670	10 Q9SP6	Q9SP6 Arabidopsis
6	207	8.6	382	16 Q9UAI5	Q9UAI5 Arabidopsis
7	197	8.2	416	17 Q9S762	Q9S762 Arabidopsis
8	187.5	7.8	388	16 Q9WZ90	Q9WZ90 Arabidopsis
9	182.5	7.6	424	5 Q9S8A2	Q9S8A2 Arabidopsis
10	181.5	7.6	381	17 Q9S512	Q9S512 Arabidopsis
11	180	7.5	412	17 Q9H000	Q9H000 Arabidopsis
12	179.5	7.5	424	5 Q9VZU8	Q9VZU8 Arabidopsis
13	172.5	7.2	370	16 Q9PEZ9	Q9PEZ9 Arabidopsis
14	169.5	7.1	381	2 Q9EVX4	Q9EVX4 Arabidopsis
15	166	6.9	382	16 Q9XW3	Q9XW3 Arabidopsis
16	165	6.9	471	3 Q8X0H8	Q8X0H8 Arabidopsis

17	163.5	6.8	360	17 Q8Z246	Q8Z246 Arabidopsis
18	163	6.8	404	16 Q8RCY0	Q8RCY0 Arabidopsis
19	163	6.8	422	16 Q8J763	Q8J763 Arabidopsis
20	162.5	6.8	385	2 Q9X4V1	Q9X4V1 Arabidopsis
21	161	6.7	380	16 Q8RA26	Q8RA26 Arabidopsis
22	161	6.7	405	10 Q9Z98	Q9Z98 Arabidopsis
23	157.5	6.6	506	3 Q13604	Q13604 Arabidopsis
24	157	6.5	511	3 Q96W6	Q96W6 Arabidopsis
25	156	6.5	416	16 Q9ZVR7	Q9ZVR7 Arabidopsis
26	155	6.5	290	2 Q8VW73	Q8VW73 Arabidopsis
27	154.5	6.4	351	17 Q58577	Q58577 Arabidopsis
28	154	6.4	358	17 Q8T2U8	Q8T2U8 Arabidopsis
29	154	6.4	422	2 Q937E1	Q937E1 Arabidopsis
30	151.5	6.3	411	17 Q26550	Q26550 Arabidopsis
31	151.5	6.3	414	16 Q9Z0S8	Q9Z0S8 Arabidopsis
32	151	6.3	375	16 Q9KC90	Q9KC90 Arabidopsis
33	150	6.3	406	16 Q8RB24	Q8RB24 Arabidopsis
34	149.5	6.2	393	16 Q9WZ95	Q9WZ95 Arabidopsis
35	149	6.2	381	16 Q8Z2F9	Q8Z2F9 Arabidopsis
36	148.5	6.2	378	16 Q9HUG1	Q9HUG1 Arabidopsis
37	148.5	6.2	383	2 Q9RHD1	Q9RHD1 Arabidopsis
38	148.5	6.2	436	16 Q9XG6	Q9XG6 Arabidopsis
39	147.5	6.2	351	16 Q8UD24	Q8UD24 Arabidopsis
40	147.5	6.2	380	2 Q8Z01	Q8Z01 Arabidopsis
41	147	6.1	333	17 Q9HSV4	Q9HSV4 Arabidopsis
42	145	6.0	395	16 Q9K6L7	Q9K6L7 Arabidopsis
43	144	6.0	409	16 Q55598	Q55598 Arabidopsis
44	144	6.0	425	2 Q939V0	Q939V0 Arabidopsis
45	144	6.0	429	16 Q8YRS3	Q8YRS3 Arabidopsis

ALIGNMENTS

RESULT 1	ID	Q9FWT0	PRELIMINARY:	PRT:	402 AA.
DT	01-MAR-2001	(TREMBLER)	16, Created		
DT	01-MAR-2001	(TREMBLER)	16, Last sequence update		
DT	01-JUN-2001	(TREMBLER)	17, Last annotation update		
DE	F1B16.5	protein.			
GN	F1B16.5				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.				
OK	NCBI_TaxID=3702;				
RN	SEQUENCE FROM N.A.				
RP	Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,				
RA	Altieri H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,				
RA	Buehler E., Chao O., Chin C., Chou J., Choi E., Gonzalez A.,				
RA	Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,				
RA	Lenz C., Liu A., Liu S., Mukharbeky N., Pham P., Sakano H., Shinn P.,				
RA	Torimaru M., Vaynsberg M., Yu G., Ecker J., Theologis A., Davis R.W.,				
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AC023754; AAC13070.1; -				
DR	InterPro; IPR001296; Glycos transf_1.				
DR	Pfam; PF00534; Glycos transf_1; 1.				
SO	SEQUENCE 402 AA; 44901 MW; 582017CDAE13D5 CRC64;				
Query Match	61.7%;	Score 1478;	DB 10;	Length 402;	
Best Local Similarity	74.5%;	Pred. No. 1.6e-107;			
Matches 283;	Conservative 53;	Mismatches 44;	Indels 0;	Gaps 0;	
Qy	89	SGGFLIMELAFILRHVGSQVWITNROSOETNDVTSLSERHMLNGVLPARGQEAVD	148		
Db	18	SRGFLIMELAFILRHVGSQVWITNROSOETNDVTSLSERHMLNGVLPARGQEAVD	77		
Qy	149	IATKADVLINRAVAGKMLDPVTKDHYKPVLPKILMTHEMGRH...VGVVHLPFVAGA	208		
Db	78	TSLKADVLINRAVAGKMLDPVTKDHYKPVLPKILMTHEMGRH...VGVVHLPFVAGA	137		

QY 209 MIDSHTAAYNSRSDRLKIQPOTYYVHLGNSKEIMEVAEDNVARVREHIRESLGV 268
DB 138 MIDSHATAGYMKRQARLGIKPKITYVHLGNSKEIMEVAEDSVAKRVLREHIRESLGV 197
QY 269 RSEDLFAIINSVSGKQDIFLAIFYALQIQHEKIKVPRIHAVVSGDVNAQTFFET 328
DB 198 RNEBLLFGIINSVSGKQDIFLAIFYALQIQHEKIKVPRIHAVVSGDVNAQTFFET 257
QY 329 QLRDFVVKNTIDRVHFVNKTIAVAPYLAIDVLYONSQSGECFCFRITTEMAFKLPVL 388
DB 258 ELNPFREKKLENFHFVNKTIAVAPYLAIDVLYONSQSGECFCFRITTEMAFKLPVL 317
QY 389 GTAGGTTEIVLDGSTGLHPAGKEGVAPLANIVRLASHAEORVSMGEKGYGRVKEMPM 448
DB 318 GTAGGTTEIVVNGTTGLHPAGKEGVAPLANIVKLATVQELRLMRGNGYERVKEMFL 377
QY 449 EHMABRIAYLKDVLKRSQ 468
DB 378 EHMABRIAYLKVLEVLQAK 397

RESULT 2

Q9FXG9 PRELIMINARY; PRT; 458 AA.
ID Q9FXG9
AC Q9FXG9; 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE F6P9.24 protein.
GN F6P9.24
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federle N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Alletti H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Bueler E., Chao Q., Chou J., Choi E., Gonzalez A.,
RA Howig B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharezy N., Pham P., Sakano H., Shlim P.,
RA Toriumi M., Vayberg M., Yu G., Becker J., Theologis A., Davis R.W.,
RA Submitted (SPP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007797; AAC12556.1;
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF00534; Glycos_transf_1;
SQ SEQUENCE 458 AA; 51367 MW; 609A1B9DBB818740 CRC64;

Query Match 61.4%; Score 1471; DB 10; Length 458;

Best Local Similarity 62.8%; Pred. No. 6, 7e-107;

Matches 304; Conservative 65; Mismatches 73; Indels 42; Gaps 6;

QY 1 MAKPSPAAVAAGRGVPHNRTOILLIVVAASASTAGFLLRQALDPCDGRG---- 56
DB 1 MAKPSTSMATLQKKRMP-----LMLLV--LSTVGMILVSTPDCSVSGRCS 50
QY 57 -----DPAALNTAVASGS--PLGFMRSKLVLSHLSISGSPILLMELAFLLRHVGSQV 110
DB 51 REKEDNSDIKIQSVGSLNPLEFMRKSLVLSHLSISGSPILLMELAFLLRGVSEV 110
QY 111 WITNQSQETNDVYSLERMLNHCQVLPARGQANDIALKQDVIYLTANTAVAKMIDPV 170
DB 111 WITNQKVEDEVIVLEHKKLDKRGVQSASQKALIDTAKSDLVVNTAVAGKMLDAV 170
QY 171 LKQHVPRVLPKILMWIHEMRGHYKPYKYLPPVAGAMIDSHTTAEYNSRSDRLKIQ 230
DB 171 LKQNVPRVLPVYLWMIHEMRGHYKPYKYLPPVAGAMIDSHTTAEYNSRSDRLKIQ 230
QY 231 MPQYVYVHLGNSKEIMEVAEDNVARVREHIRESLGVRSDDLFAIINSVSGKQDIF 290
DB 231 MPKTYVYVHLGNSKEIMEVAEDSVAKVLR-----VSRGKQDLF 269

QY 291 LQAFYALQIQH-EKIKVPRIHAVVSGDVNAQTFFETQLRDFVVKNTIDRVHFVNKT 349
DB 270 LRFHESLKIYKEIKLEVPYTHAVVSGDVNAQTFFETELNPFQEMKQOKIVHVNKT 329
QY 350 LAVAPYLAIDVLYONSQSGECFCFRITTEMAFKLPVLGTAGGTTEIVLDGSTGLHP 409
DB 330 MKVAPYLAIDVLYONSQSGECFCFRITTEMAFKLPVLGTAGGTTEIVVNTTGLHN 389
QY 410 AKKEGVAPLANIVRLASHAEORVSMGEKGYGRVKEMPMFHMABRIAYLKDVLKRSQ 469
DB 390 TKGDDVLPLANKIVKLATVQELRLMRGNGYERVKEMFL EHMABRIAYLREVLQAKI 449
QY 470 HSRP 473
DB 450 HSRP 453

RESULT 3

O9LI48 PRELIMINARY; PRT; 188 AA.
ID O9LI48
AC O9LI48; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EST C74729 (E50675) corresponds to a region of the predicted gene.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriactoidae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Macdonato T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: p0469B09."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP01366; BAA92419.1;
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF00534; Glycos_transf_1;
SQ SEQUENCE 188 AA; 20622 MW; C7C4E168752107DA CRC64;

Query Match 28.0%; Score 671; DB 10; Length 188;

Best Local Similarity 91.6%; Pred. No. 8, 7e-45;

Matches 131; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 311 IHAVVSGDVNAQTFFETQLRDFVVKNTIDRVHFVNKTIAVAPYLAIDVLYONSQSG 370
DB 1 MHAVVSGDVNAQTFFETQLRDFVVKNTIDRVHFVNKTIAVAPYLAIDVLYONSQARG 60
QY 371 ECGRRTTEMAFKLPVLGTAGGTTEIVLDGSTGLHPAGKEGVAPLANIVRLASHAE 430
DB 61 ECGRRTTEMAFKLPVLGTAGGTTEIVLDGSTGLHPAGKEGVAPLANIVRLASHAE 120
QY 431 QRVSMGEKGYGRVKEMPMFHMABRIAYLKDVLKRSQ 453
DB 121 DRVSMGRKGYGRVKEMPMFHMABRIAYLKDVLKRSQ 143

RESULT 4

O9LSB5 PRELIMINARY; PRT; 697 AA.
ID O9LSB5
AC O9LSB5; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Glycoyl transferases-like protein (AT3G15940/MYC8_7).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RL Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RN Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 [3]
 RN SEQUENCE FROM N.A.
 RP Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones.";
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AB026653; BAB02880.1; -;
 DR EMBL; AY091763; AAM10311.1; -;
 DR InterPro; IPR001199; Cyt B5.
 DR InterPro; IPR001296; Glycos transf_1.
 DR Pfam; PF00534; Glycos transf_1; 1.
 DR PROSITE; PS00191; CYTOCHROME_B5_1; UNKNOWN_1.
 KW Transfease
 SQ SEQUENCE 697 AA; 78706 MW; 17312545669F315 CRC64;
 Query Match 20.3%; Score 487; DB 10; Length 697;
 Best Local Similarity 29.1%; Pred. No. 1.7e-29;
 Matches 147; Conservative 77; Mismatches 159; Indels 122; Gaps 13;
 QY 52 CDGGRDPAALNTAAVAGSSPLGPMR---SKVLVLSHLSLGGPPLLEAFLLRHVQ 108
 DB 223 CDRKSD-----FRLVWSRRFVLPHLSMTGAPLSMELASSELLSCGAT 267
 QY 109 VVWITNQRSGETNDVTVSLERMLNHGVQVLPARGOAVDIALKADVLINTAVAGKMD 168
 DB 268 VVAVLSRRG-----GLGDELTRRIIVYEDKGLSFTAKKADIVLAGSNVCASWD 320
 QY 169 PVLKQVPRVLPKILMITHMRGRHYK----VEVYKHLFPVAGAMISHTTAAYNNS R 222
 DB 321 QYM-DHPRAGGSQIAMWMMNRREYDRAKPVLRVLLIFL-----SEVQSKMLTWC 373
 QY 223 TSDRLKIOMPQTVVHVGNSKEIMEVAEDNVA-----RVLRHEHRESLAV 268
 DB 374 EEDHVYKLR-SQPVIVPLSVDELAFVAGVSSSLMTPLTQETWKEKRGKLESVRTFGL 432
 QY 269 RASDILLFALINSVSRGKQDLFLQATYQALDL-----IQHKLK 307
 DB 433 TDDKMLVMSLSINSINGQLLLLESVALLEREQTOBOVAKRNOSKIIKNINGIRKSKIS 492
 QY 308 VPRIHAV-----VVGSDVY 321
 DB 493 LSAHRHLRSGSRKMKITS PAVDNHPSVLSATGRKRLILSGNVTQKDKLLLSVGSKSN 552
 QY 322 AQCKETQULDRFVYKN-TIHDRYHFNKTLNAPLALIVLVNQSQRGRCFRTTTE 380
 DB 553 -KVAAYVKEMLSPLSNNGNLNSVLMPTATRTVASYSAADVYTNNSQVGSTFRFVITE 611
 QY 381 MAFKLVLGTAAAGTTEIVLDGSTGLHVPAGKEGVAPLAQNIYRLASHAEORVSMGKGY 440
 DB 612 MAYGLPVLGTDAAGTKEIVEHANTVGLHVPBAGNKVLAQNLLFLLRNPSRRLDLSQGR 671
 QY 441 GRVKEFMENHRAEKLAAVLKQVLR 465

DB 672 ELVEKMTWKQHMRYKRPVDVLYKCMR 696

RESULT 5

Q9SSP6 PRELIMINARY; PRT; 670 AA.

AC Q9SSP6;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE F6D8.36 protein (Glycosyl transferase, putative).

GN F6D8.36 OR F19K6.13.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Euphorbia; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC euroside II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,

RA Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Mukharasy N., Sakano H.,

RA Vaysberg M., Chin C., Choi E., Chou J., Altati H., Araujo R.,

RA Brooks S., Buehler E., Chao Q., Conn L., Conway A.B., Dunn P.,

RA Hansen N., Hwang B., Huizar L., Khan S., Kim C., Palm C., Rowley D.,

RA Shinn P., Walker M., Davis R.W., Ecker J.R., Federspiel N.A.,

RA Theologis A.;

RT "Arabidopsis thaliana chromosome 1 BAC F6D8 sequence.";

RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

RA Dunn P., Egu P., Feldblyum T.V., Peng J.-D., Peng B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremetskaia I., Kritz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.-P.,

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Martelli A.,

RA Millichaer J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,

RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,

RA Utracki T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

RA Wu D., Yu G., Fraser C.W., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

thaliana.";

RT Nature 408:816-820(2000).

DR EMBL; AC008016; AAG55621.1; -;

DR EMBL; AC037424; AAG51540.1; -;

DR InterPro; IPR001199; Cyt B5.

DR InterPro; IPR001296; Glycos transf_1.

DR Pfam; PF00534; Glycos transf_1; 1.

DR PROSITE; PS00191; CYTOCHROME_B5_1; UNKNOWN_1.

KW Transfease.

SQ SEQUENCE 670 AA; 75636 MW; F59FE275C33DD0AF CRC64;

Query Match 20.1%; Score 482.5; DB 10; Length 670;

Best Local Similarity 29.2%; Pred. No. 3.6e-29;

Matches 148; Conservative 77; Mismatches 173; Indels 109; Gaps 13;

QY 38 STAGFLLR--GALRD-----PCDGRDPAALNTAAVAGSSPLGPMR---SKVLV 80
 DB 193 ATYGKLLDPFSSLEDEVLKWSPHRRSGTCDKSD-----FRLVWSRRFV 237
 QY 81 IVSHSLSSGGPPLLEAFLLRHVGSQVWITNQRSGETNDVTVSLERMLNHGVQVLP 140
 DB 238 LFLHLSMTGAPLSMELASSELLSCGATVSAADVYTNNSQVGSTFRFVITE 290

QY 141 ARGDAVDIALKADLVILNTAVAGKMDPVLKDVHPKILWIIHEKRGHYFK----- 195
DB 291 DKGEISFRTAMKADLLIAGSAVCTSWIDQYMNHH-PAGSGQIAMIWENREYFDRAPV 349
QY 196 VEVYVCH.PFVAGAMIDSHHTAEVYNN--RTSDRLKIOMPTVYVHLGNSKELMEV----- 249
DB 350 LDRVXKMLFL-----SESQSRQMLTWCEEEHFKLA-SQPIYVPLSVNDELAFVGISS 402
QY 250 -----EDNVARVLEIRHSICVSEDLIPAIINSVRKGGDLFLQAFYQALQ- 299
DB 403 LNTPLSPERQVRKRIRESVTELGITSDMLVMSLSINPTKGLLIESIALALSE 462
QY 300 -----LTHQKLVPRIRHAY-----VYGS 318
DB 463 RQGSQRNNHGIIIRKKNVSLSKRLRGSSRQMKSVSLTDLNGIRREKQELKVLGSGVS 522
QY 319 DVNAQTFFETQLDFPVVKNTHDRVHFNKTLAVAPYLAIDVLVONSQGRCEGRTTI 378
DB 523 KSNKGVYKEMLSFLSNSGNLSKSVMTPTATTVASLVSADVYVTVNSQVGETGRTYI 582
QY 379 EAMAFKLPVLGTAGTTEIVLDGSTGLHPACKEGVAPLAKNIVRLASHAQRVSMGEK 438
DB 583 EAMAVGLAVGTDTAGGTKEWQHMTGLTSMGRSGNKLHNLVLLRNPDRRLQSE 642
QY 439 GYGVKEMEMEHMAERLAAYLKDVLR 465
DB 643 GRKMYEKYMKQHMVYKRFVDVLVYKCMR 669

RESULT 6

080U15 PRELIMINARY; PRT; 382 AA.

AC 080U15, Setubal J.C., Kaul R., Monke D.E., Kitajima J.P.,
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Glycosyltransferase.
GN ATU3560 OR AGR L.2541.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kitajima J.P.,
RA Okura Y., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Faulstich J., Eisen J.A., Kaip P.D., Bove D., Str.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.,
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Curcio B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Houtrel K., Gordon J., Vaudin M., Jarchock O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markfeld B.,
RA Flanagan C., Crowell C., Gurson J., Iomo C., Sear C., Strub G.,
RA Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
DR EMBL; AE009285; AA44372.1; -
DR EMBL; AE008327; AA89838.1; -
KW Transferase; Complete proteome.

SQ SEQUENCE 382 AA; 40400 MW; 802885D24D0F042A CRC64;
Query Match 8.6%; Score 207; DB 16; Length 382;
Best Local Similarity 27.5%; Pred. No. 5,9e-08;
Matches 68; Conservative 46; Mismatches 109; Indels 24; Gaps 7;
QY 220 NSTRSRLKI---OMPTVYVHLGNSKELMEVADNVARRVIREHRSIGVSEDLIF 275
DB 154 NSERTGAPTEGEGADRVIRVYNGFDPAKAKLHDAGMAAR-----LRNKGILPQPLV- 207
QY 276 AIINSVRKGGDLFLQAFYQALQLOHFKLVPRIRHAYVNSDVNAQCKEPTOLRPFVV 335
DB 208 GLFRLESEKQGHVFLDL--AAHEGVQ-----AVVGAALFGQEAFAIRRQAS 256
QY 336 KNTHDRVHFNKTLAVAPYLAIDVLVONSQGRCEGRTTIEMAFKLPVLGTAGT 395
DB 257 FLGIDRVRFVFLGFRSDVPELMASMDVVAHTSI-VAEPFGVVEAMCGRPVAVTRGGGV 315
QY 396 TEIVLDGSTGLHPACKEGVAPLAKNIVRLASHAQRVSMGEKGYGVKEMEMEHMAER 455
DB 316 TEITRGEITGLVPPG--DASALAAITLISDPAKRLGSGREDVSDRFSIQETCRS 373
QY 456 IAYLVLD 462
DB 374 VSALLTE 380

RESULT 7

058762 PRELIMINARY; PRT; 416 AA.

AC 058762, Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein PH1035.
GN PH1035.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Hatake Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohtoku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshitawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76 (1998)
DR EMBL; AF000004; BA30133.1; -
DR InterPro; IPR001296; Glycos. transf. 1.
DR Pfam; PF00534; Glycos. transf. 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 416 AA; 48196 MW; 89BDBE51EC7B0 CRC64;

Query Match 8.2%; Score 197; DB 17; Length 416;

Best Local Similarity 25.5%; Pred. No. 4.1e-07; Indels 74; Gaps 18;
Matches 104; Conservative 56; Mismatches 174;

QY 87 SLSGPFL-LTMEFLRLRHVGSQVWITNRSQRTNDVYVLSLEHMLHNGVQLPARQGE 145
DB 50 SFGGVAELTSLVFLRISGTEARWFIETGPEFFVYVTKTHNL-----QGNB 99
QY 146 AVIADIAKADVLIN-TAVAGKMDPVLKDVHPKILWIIHEKRGHYFNVYVYKLPF 204
DB 100 SLGLTEMKELYLNVNRENSKFTIDSSFDV-----LVHDDQPALILEFEYKKSFW 150
QY 205 VAGAMID-SHTTAEVYNN-----SRTSDRLKIOMPTVYVHLGNSKELM-----EVAEDN 252

Db 151 LMRCHIDSSPNEFWEFLRRFVEKYORVYI.FHLPEYVQPELDRNKAVIMPESIDPLSEKN 210
Qy 253 VARR---VLREHRESIGRSDDLFAINSYVR---GGGDLFLQAFYALOLIHETL 306
Db 211 VELKOTELLR--LIERPDVPEK---PITTVGSPFPMWG-----IFVDITRYKVE 258
Qy 307 KVPRIHAVVVG---SDVNAQTFEFTOLR---DPVVK---NTIHDRVHPVUKTLAAVRY 355
Db 259 KIPGQOLLVGVMAHDDPEGWITFPEKTLRKIGEDYDVKLTWLIQVHAREVN-----AF 312
Qy 356 LMAIDVYONGSGEGEFGFTITTEMAFKLPVGTAGTTEI.VLDGSTGLHPAGKEGV 415
Db 313 QRSRVILQMSIREG--FGLTYTEMWKGKIVGRVAGIKKQIYDGETGVLRDANENV 370
Qy 416 APLAKNIYRLASHAEOVSNGEKGGRVYKEMEMHMAERIALAVKDV 463
Db 371 ----EKVLYLKHPREVSKEWGAKEKRYKKNFITTKEHMERIDIINSL 414

RESULT 8

Q9WZ90 PRELIMINARY; PRT; 388 AA.
AC Q9WZ90;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Lipopolysaccharide biosynthesis protein, putative.
GN TM0622.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.U.,
Hate D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt W.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
R. Salberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001736; AAD35706.1; -
DR TIGR; TM0622; -
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Complete proteome.
SQ SEQUENCE 388 AA; 43648 MW; 6BA1BE01157F306 CRC64;

Query Match

Best Local Similarity 7.8%; Score 187.5; DB 16; Length 388;
Best Local Similarity 23.5%; Pred. No. 2e-06;
Matches 88; Conservative 58; Mismatches 136; Indels 93; Gaps 13;
Qy 136 VQVLPARGOAVDIALKADVLVLTAVAGMIDPVLKHKVLPKILMIHMGHFK 195
Db 47 VEKLTSGQYVTVIYDYE-----AIAPS-----KVIRLLRLAIKMERITNL 89
Qy 196 VEYVK-----HLFVAGAMIDS-----HTTAE-----YVNSKTSDRDKI 229
Db 90 LREIRPDILHSHLSARIALIPALLCRIIPKVHTIHTVAEKDAKIGTRFPN-----RI 142
Qy 230 QMPQTVVHLGNSKELMEVAENNAKRVLAENHRESIGV-----RSEDLLFAI 277
Db 143 AKFEGFVPVISIOEVAESVKLYGKIKSTPVLYNGIDQKFSIDQKRVDRKILLNV 202
Qy 278 INSVSRGKGDVFLQAFYALOLIHETLKVPRIHAAVVGSDVNAQTFEFTOLRDFVKN 337
Db 203 AR-LSREKNALLVRAFSKAVQ-----SCPMLIELMLVGDG-----ELRRDIEELVQOL 249
Qy 338 TIHDRVHPVUKTLAAVRYLAIDVYONGSGEGEFGFTITTEMAFKLPVGTAGTTE 397

Db 250 GLEEKVFEQVSDVPELLSQADIFVLSQDYG--FGLVAMAMAGLPVITATGIGPE 307
Qy 398 IVDGSTGLHPAGKEGVAPLAKNIYRLASHAEOVSNGEKGGRVYKEMEMHMAERIA 457
Db 308 ILEGGRAGLIVP--PROVDALAKAIVELARDEKRAELSDYGRGLVAERF----- 355
Qy 458 AVLKDVLRKQSEHSR 472
Db 356 ----DIRRTVREYER 366

RESULT 9

Q95SA2 PRELIMINARY; PRT; 424 AA.
AC Q95SA2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE GM04690P.
GN CG1291.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda C.J.,
RA Munoz J., Pacheb J., Paragas V., Park S., Phouanenvong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;
RL Submitted (JCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AT060893; AAL28441.1; -
DR FlyBase; FBgn0035401; CG1291.
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
SQ SEQUENCE 424 AA; 48639 MW; F39F08BF7A11B35 CRC64;

Query Match

Best Local Similarity 7.6%; Score 182.5; DB 5; Length 424;
Best Local Similarity 22.2%; Pred. No. 5.8e-06;
Matches 101; Conservative 75; Mismatches 179; Indels 99; Gaps 15;
Qy 79 VLVSHLSLGGPULMELAFLLRHVSQVVTNQRS-----QETNDVYSIHERMLN 133
Db 4 VLFHPDLGIGABRLVVDAAALKEKHQVSPLTNHDSTHCFKELADGTFV----- 57
Qy 134 HGVQVLPARGOAVDIALKADVLVLTAVAGM-----LDPVLKDHVPKVLKILMI 186
Db 58 HVGQDLPRGLGEGRFYALICVYLMYLAIVASFPMQREQVDVVCGLISVCP-----V 112
Qy 187 HEKRGHYFVEYVVELPVGAMIDSHHTAEVNSKTSDRDKIOMPQTVVHLG-----N 241
Db 113 LRAPRPVPLVFCYCFP-----DQLSSRBGLKRLPLPIMLEBHTGLADKVLVN 165
Qy 242 SKELMEVAEDNVAR-----RVLR-----EHRESIGVRSDDL-----LF 275
Db 166 SKFTLVFPQDTRRLSTVVDVLYSLHTGYFPOKKEORSALLDEPVHPRVLANFIY 225
Qy 276 AINSVSRGKGDVFLQAFYALOLIHETLKVPRIHAAVVG-----SDVNA 322
Db 226 LDINRERKKNALHSLRLLDGMLPATYFK--RCRLIIAGYDTRCMENVEHFALEH 283
Qy 323 QTFEFTOLRDFVKN-----KNTIHDRVHPVUKTLAAVRYLAIDVYONGSGEGEFGFT 377
Db 284 LTE-ELKQDHHVLLRSPTDEKCRLLFAAHCLLYTP-----ENEHFGIVP 328
Qy 378 IEAARFKLPVLTAGTTEI.VLDGSTGLHPAGKEGVAPLA* IYRLASHAEOVSNGE 437
Db 329 LEGYTSKRPVLAINGSGTETVYVNTSTGFLCEKTEKS---F*AMQLRDRDQLRVKXGD 385
Qy 438 KQYGRVYKEMEMHMAERIAAVLKDVLRKQSEHS 471

Db 386 QGHRVQKFSFQAFADRLNGIIRDLVPISRSS 419

RESULT 10

059512 PRELIMINARY; PRT; 381 AA.

AC 059512; 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Hypothetical protein PH1844.

GN PH1844.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OK NCBI_TaxId=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=9834437; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohtoku Y.,

RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki Y., Kushiida N., Oguchi A.,

RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

RA Menuchi Y., Shizuya H., Kikuchi H.,

RT "Complete sequence and gene organization of the genome of a hyper-

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."

RL DNA Res. 5:55-76(1998).

DR EMBL; AF000007; BAA30965.1; "

DR InterPro; IPR001296; Glycos transf.1.

DR Pfam; PF00534; Glycos transf.1; 1.

DR Hypothetical protein; Complete proteome.

KW SEQUENCE 381 AA; 42826 MW; EDOBOCA0CFSCAE1 CRC64;

Query Match

Best Local Similarity 7.6%; Score 181.5; DB 17; Length 381;

Matches 93; Conservative 67; Mismatches 161; Indels 81; Gaps 15;

RESULT 11

09HH00 PRELIMINARY; PRT; 412 AA.

AC 09HH00; 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Putative trehalose synthase.

GN PF1742.

OS Pyrococcus furiosus, and

OS Thermococcus littoralis.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OK NCBI_TaxId=2261, 2265;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=P. furiosus, and T. littoralis;

RX MEDLINE=20566786; PubMed=1115105;

RA Diruggiero J., Dunn D., Maeder D.L., Holley-Shanks R., Chataud J.,

RA Horlacher R., Robb F.T., Boos W., Weiss R.B.,

RT "Evidence of recent lateral gene transfer among hyperthermophilic

RT archaea."

RL Mol. Microbiol. 38:684-693(2000).

DR EMBL; AF010272; AAL81866.1; "

DR EMBL; AF307053; AAG45391.1; "

DR EMBL; AF307052; AAG45375.1; "

DR InterPro; IPR001296; Glycos transf.1.

DR Pfam; PF00534; Glycos transf.1; 1.

DR Hypothetical protein; Complete proteome.

KW SEQUENCE 412 AA; 48017 MW; 6A5338298A6B5CEA CRC64;

Query Match

Best Local Similarity 7.5%; Score 180; DB 17; Length 412;

Matches 99; Conservative 55; Mismatches 158; Indels 112; Gaps 18;

RESULT 12

09VZ08 PRELIMINARY; PRT; 424 AA.

AC 09VZ08; 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Putative trehalose synthase.

GN PF1742.

OS Pyrococcus furiosus, and

OS Thermococcus littoralis.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OK NCBI_TaxId=2261, 2265;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=P. furiosus, and T. littoralis;

RX MEDLINE=20566786; PubMed=1115105;

RA Diruggiero J., Dunn D., Maeder D.L., Holley-Shanks R., Chataud J.,

RA Horlacher R., Robb F.T., Boos W., Weiss R.B.,

RT "Evidence of recent lateral gene transfer among hyperthermophilic

RT archaea."

RL Mol. Microbiol. 38:684-693(2000).

DR EMBL; AF010272; AAL81866.1; "

DR EMBL; AF307053; AAG45391.1; "

DR EMBL; AF307052; AAG45375.1; "

DR InterPro; IPR001296; Glycos transf.1.

DR Pfam; PF00534; Glycos transf.1; 1.

DR Hypothetical protein; Complete proteome.

KW SEQUENCE 412 AA; 48017 MW; 6A5338298A6B5CEA CRC64;

Query Match

Best Local Similarity 7.5%; Score 180; DB 17; Length 412;

Matches 99; Conservative 55; Mismatches 158; Indels 112; Gaps 18;

DT 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE CG1291 protein (R667594p).
 CN CG1291
 OS Drosophila melanogaster (Fruit fly)
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan U.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
 RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gottlieb J.H., Gu Z., Guan P., Harris W.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaisi M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA LaRoche J.E., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollack J., Puri V., Reese M.G.,
 RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhu G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N. A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnak D., Farfan D., Fiske E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Munzall C.J., Nuncio J., Pacle J., Pargava V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RX EMBL: ARO03477; AAF47219.1; -;
 DR EMBL: AY071629; AAL49251.1; -;
 DR FLYBase; FBgn003401; CG1291.
 DR InterPro; IPR001296; Glycos transf. 1.
 DR Pfam; PF00534; Glycos transf. 1; 1.
 SQ SEQUENCE 424 AA; 48598 MW; 9F77DA4610DAB34 CRC64;

Query Match 7.5%; Score 179.5; DB 5; Length 424;
 Best Local Similarity 22.2%; Pred. No. 9.9e-06;
 Matches 101; Conservative 73; Mismatches 181; Indels 99; Gaps 15;

QY 79 VLVSVHSLSSGGPRLIMELAFLLRHVGSQVWITNORS-----QETNDVYLSLRRMLN 133
 DB 4 VLPFLPRLGIGSGBRLVVDALALAKRGHVSLTNHDSHCKETADGTPTV----- 57
 QY 134 HGQVULVARGQEVNDIALKADLVLTNTAVAGK-----LDPLKQHVVKLPKILWMI 186
 DB 58 HVGQDMLPRGLFGRFYALCATLRMLVAIYASFMPQREQDVVVCOLSVCP-----V 112
 QY 187 HEMRGHVFKEVYKHLPFVAGAMIDSHTTAEWNSRTSDRIKIQMPQTVYHLG-----N 241
 DB 113 LRFAPHRPKLFCYCHP-----DQLSSRGGLKRLYLRLPIWMEHNTIGLADKVLVN 165
 QY 242 SKELMEVAENNVAR-----RVLK-----EHRESLQVRSSELD-----LF 275
 DB 166 SFTLRVFDQTFRRSTVDPVLPVLSLHTQYFDQMDKLEQKRSALDSEVPVPLNAPLY 225
 QY 276 ALINVSRGQDQLFQAFQYQALQILQHEKLPKPRHIAVVVG-----SDVNA 322
 DB 226 LDIVRYERKKNHIALHSLRLGDMPLPATDFK--RCRLITAGVDTRCMEVNHFALEH 283
 QY 323 QTKPEQLQRPFVY-----KNTHRVHFVVKTLVAVPILALDVLVONSQSGEGEGRIT 377
 DB 284 LTF-ELKIQHVVLLSPTEDEKCRLLFAHCLLYTP-----ENEHGIVP 328
 QY 378 IEAAAFKLPVLTAGCTTEIVLDSTGLHPAGKGVAPLAKNYVLSHAEQVRSNGE 437
 DB 329 LRGWCKSPVVALNSGPTFETVSTGFLCEKTEKS---FGAMQLQFPRDGLRVKMGD 385
 QY 438 KQYGVEMEMEHMARIANLVKDVLRKQEH 471
 DB 386 QGHRKVOOKFSFOAFADRLNGILRDVLPISSRES 419
 RESULT 13
 ID 09PEZ9 PRELIMINARY; PRT; 370 AA.
 AC 09PEZ9;
 DT 01-OCT-2000 (TRENBLREL. 15, Created)
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE Lipopolysaccharide biosynthesis protein.
 CN XF0879.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OX NCBI_TaxId=2371;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365177; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barroes M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Gantier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kurnae E.E., Laigret F., Lambis M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado W.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteleiro-Vicorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Rodrigues V.B., Pasqueto J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V.B., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Slyke M.A., Verjovskij-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Secubal J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE003927; AAF83689.1; -
 DR InterPro: IPR001296; Glycoe_transf_1.
 DR Pfam: PF00534; Glycoe_transf_1; 1.
 KM Complete proteome.
 SQ SEQUENCE 370 AA; 41001 MW; 57E3151F08B85510 CRC64;

Query Match 7.2%; Score 172.5; DB 16; Length 370;
 Best Local Similarity 25.0%; Pred. No. 2.9e-05;
 Matches 79; Conservative 48; Mismatches 136; Indels 53; Gaps 12;

QY 178 VLPRIL-----WTHMRGHYFKYV--KG-LPFGAGMIDSHTEYNASTSR- 226
 DB 78 VHPDVLGDFGKRYPLMLMGRLYRLFVLFHRLP-----PNCSTYVPPRLADRF 130
 QY 227 -----LKIQPTYYVHLGNSKELMEVAD-NVARRVLRHRESIGVRSDDL 274
 DB 131 FAVSAYARRHYLAEGMPPEVQVLXNPVTDALRPDPVRRLMHE-----LGMDEDVLY 185
 QY 275 FAIINSVSGKODLFLQAFYQALQLQHEKLVPRHNAVVGSDVNAQTKREYQADPV 334
 DB 186 VGCGRHSGKGVFLAEPMQAMQ-----EEPRLCCLMMGTGLHVO-----RLAATV 233
 QY 335 VKNTIHDRVFNKTLAVAPYLAIDVLYVONSQSGEGCFGRITTEMAKFLPVLTAGG 394
 DB 234 AGSFAPASRGVGLWTDPARFYQALMTAMPST-LPFTFGVSAEAGVGLVSDVGG 292
 QY 395 TTEVLDSGTGLHPAGKEGVAFLAKNIVRLASHAEQVSGKEGGRKEMFHHMAE 454
 DB 293 AAEVLQDGTITMLPRGD---VPAMRYALACPDQPPRAAMAAAPSVLEAFNSQVYTA 349
 QY 455 RIAAVLKDYLRKSQER 470
 DB 350 EFISELERVLI--SDRH 363

RESULT 14

O8EXK4 PRELIMINARY; PRT; 381 AA.
 AC Q9EXV4;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Putative hexose transferase.
 GN CPGC.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20519245; PubMed=11065358;
 RA Altmann-Reig E., Mulholland F., Gasson M.J., Griffin A.M.,
 RT "The complete cps gene cluster from *Streptococcus thermophilus* NC8B
 RT 2393 involved in the biosynthesis of a new exopolysaccharide,"
 RL Microbiology 146:2793-2802(2000).
 DR EMBL: Y17900; CAC18357.1; -
 DR InterPro: IPR001296; Glycoe_transf_1.
 DR Pfam: PF00534; Glycoe_transf_1; 1.
 KM Transferase.
 SQ SEQUENCE 381 AA; 42630 MW; 995210276BD179AC CRC64;

Query Match 7.1%; Score 169.5; DB 2; Length 381;
 Best Local Similarity 25.5%; Pred. No. 5.1e-05;
 Matches 76; Conservative 53; Mismatches 124; Indels 45; Gaps 14;
 QY 160 TAV-AGKMLDPVLDKHPVKVLPKILMWIHEMGHYFKYEVK--HLDPVAGAMIDSHTAA 217
 DB 104 TAVLEGIVLKRLK-----LP-LIHVHVEI---IVKPAISDFINFLMGRAVDITV- 152

QY 218 YNNSRTSDRLK-----IQPQTYVHLGNSKELMEVADNVARRVLRHRESIGVRSDDL 273
 DB 153 --SNAVAHNVKQSFVKNDQVITYNG-----VDNAVHEMNASAVRDOFGIVODAL 202
 QY 274 LPAIINSVSGKODLFLQAFYQALQLQHEKLVPRHNAVVGSDVNAQTKREYQADPV 332
 DB 203 VIGMVRVNAKMGQGNFLNVTPLK-----ANPKAVAFLAGAFEGEWRVDELETA 255
 QY 330 LRDVFNKTIHDRVFNKTLAVAPYLAIDVLYVONSQSGEGCFGRITTEMAKFLPVLT 389
 DB 256 ISDPVAGQI-KRIDYKSTTEL--YNNMFDIVLPST-NPDPFLPVLEAMCGKPVVGG 310
 QY 390 TAGGRTTEVLDSGTGLHPAGKEGVAFLAKNIVRLASHAEQVSGKEGGRKEMFHHMAE 447
 DB 311 YRIGGVCEVKEGNGILL--ATPNQPAELSKATIELMDITERKEDFGKASVKNOKELF 366

RESULT 15

O8YOM3 PRELIMINARY; PRT; 382 AA.
 AC O8YOM3;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical protein A1R3699.
 GN A1R3699.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Itiguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsunoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium *Anabaena* sp. strain PCC 7120,"
 RL DNA Res. 8:205-213(2001).
 DR EMBL: AF003594; BAB75398.1; -
 DR InterPro: IPR001296; Glycoe_transf_1.
 DR Pfam: PF00534; Glycoe_transf_1; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 382 AA; 41501 MW; 2D7546639F2A7B67 CRC64;

Query Match 6.9%; Score 166; DB 16; Length 382;
 Best Local Similarity 20.9%; Pred. No. 9.7e-05;
 Matches 76; Conservative 75; Mismatches 144; Indels 68; Gaps 13;

QY 104 HVGSOYVWITNORSQETNDVYSLERMLNNGVYLPRAGQ-----EAVDIALKADLV 157
 DB 48 HIRVEV--FTNQ-----PIQVRKQSNLQAFSGQLADPLAVAKVQTAHEVDL 95
 QY 158 LNFVAVAKMLDPVLDKHPVKVLPK--LIWTHIHE-MRGHYFKYEVKHLPPYAGAMIDSHT 214
 DB 96 ANQYK-----LVGKIASFIARRPVYHLDLSPEHFSQTNLR-----VAYNLNRA 145
 QY 215 TAEYNSRTSDRLKIQ--MPQTYVHLGNSKELMEVADNVARRVLRHRESIGVRS 270
 DB 146 SLVIANQSQSTAFIQGGRAEITVYINGFDINLYTKTSPDISK-----LQOQGV-A 198
 QY 271 EDLLFAIINSVSGKODLFLQAFYQALQLQHEKLVPRHNAVVGSDVNAQTKREYQADPV 330
 DB 199 NNFVYGHFSRLSPWKGHILIDLAQC-----PQVYALIVGDALFGODYVKEL 248
 QY 331 RDFVFNKTIHDRVFNKTLAVAPYLAIDVLYVONSQSGEGCFGRITTEMAKFLPVLT 390
 DB 249 HQQITRLGLENRKVELEGFRADIPQLMAACD-LVAHTSTAPPEFGHVIEMALCGKPVVAA 307
 QY 391 AAGGTEVLDSGTGLHPAGK-----EGVAPLAKNIVRLASHAEQVSG 435

Tue Jul 8 09:50:08 2003

us-09-938-294-45.rpt

Page 9

Db	308 KAGAMELVEHGVNGFLTPGSECELANINTCIEDTOKTATIASNAQAIASQRFVVTI	367
Oy	436 GEX 438	
	:	
Db	368 NOQ 370	

Search completed: June 30, 2003, 16:27:36
Job time : 41 secs
